



```

QY 241 NMKVEEYVWKEHRELDRTGRKGIYIKGTSERLTMLVEHSHVYDPTFIEDLL 300
|
|
|
Db 241 NMKVEEYVWKEHRELDRTGRKGIYIKGTSERLTMLVEHSHVYDPTFIEDLL 300
QY 301 TYRTFLSSPMEVGKLLLEWNPDSLRDKYTRVLLVMVNNHNFDEGDPAMTFLEEFENN 360
|
|
|
Db 301 TYRTFLSSPMEVGKLLLEWNPDSLRDKYTRVLLVMVNNHNFDEGDPAMTFLEEFENN 360
QY 361 LEREMGHLRLNTACAKAKARRLMTLTKPSREAPLPFILLGSEKGFIVDSVDSGS 420
|
|
|
Db 361 LEREMGHLRLNTACAKAKARRLMTLTKPSREAPLPFILLGSEKGFIVDSVDSGS 420
QY 421 KATEAGLKRQDQILEVNGCFENFIOLSKAMELIRNNTHTSTVKNLVEFKELRLSEE 480
|
|
|
Db 421 KATEAGLKRQDQILEVNGCFENFIOLSKAMELIRNNTHTSTVKNLVEFKELRLSEE 480
QY 481 KNGAPHLPKIGDIKASRYSTIPDLAVDEQYIGLEKYNKSKANTVGRNKKKILDKT 540
|
|
|
Db 481 KNGAPHLPKIGDIKASRYSTIPDLAVDEQYIGLEKYNKSKANTVGRNKKKILDKT 540
QY 541 RSLIPKPYNDIGIGSODDSIVGLRQTKHPTALPVSGTSSNPDLQSHHRLDFS 600
|
|
|
Db 541 RSLIPKPYNDIGIGSODDSIVGLRQTKHPTALPVSGTSSNPDLQSHHRLDFS 600
QY 601 ATPDLPDVLRVFKADQDSRYIMISKDTAKEVYIOAREFAVTPPOYSICEVSYTPE 660
|
|
|
Db 601 ATPDLPDVLRVFKADQDSRYIMISKDTAKEVYIOAREFAVTPPOYSICEVSYTPE 660
QY 661 GVIKORRLPDOLSKLADRIQLSGRYLYLKNMETETLCSDEDAQELRESQSLQLSTVE 720
|
|
|
Db 661 GVIKORRLPDOLSKLADRIQLSGRYLYLKNMETETLCSDEDAQELRESQSLQLSTVE 720
QY 721 VATOLSMNFELFRNIEPTETIDOLFRLKSTSCANLKRFEVINOEFVWASELRETN 780
|
|
|
Db 721 VATOLSMNFELFRNIEPTETIDOLFRLKSTSCANLKRFEVINOEFVWASELRETN 780
QY 781 QLRKMKIKHFKIALHCRECKNFNSMFAIISGLNLAVARLRTTWELPKRYEKLFDL 840
|
|
|
Db 781 QLRKMKIKHFKIALHCRECKNFNSMFAIISGLNLAVARLRTTWELPKRYEKLFDL 840
QY 841 ODLDPSPNMAKRYNVLSQNLQPPITLPFVYIKDLTFLEHGNDSKVDGLVNEFKLRMI 900
|
|
|
Db 841 ODLDPSPNMAKRYNVLSQNLQPPITLPFVYIKDLTFLEHGNDSKVDGLVNEFKLRMI 900
QY 901 AKETIRHGRMASVMDPALMFRTRKKKWRSLGSLSGSTNATVLDVAQTGGHKKRVRSS 960
|
|
|
Db 901 AKETIRHGRMASVMDPALMFRTRKKKWRSLGSLSGSTNATVLDVAQTGGHKKRVRSS 960
QY 961 FLNAKKLYEDAQARKVQYLSNLELMEDESLQTLSCGEPATNTLPKNGDCKPVKSE 1020
|
|
|
Db 961 FLNAKKLYEDAQARKVQYLSNLELMEDESLQTLSCGEPATNTLPKNGDCKPVKSE 1020
QY 1021 TSPVAPRAGSQQAQSLPQOQOQPPRAHKINOGIQVAVSLYPSRKKVPVKDLEPGINS 1080
|
|
|
Db 1021 TSPVAPRAGSQQAQSLPQOQOQPPRAHKINOGIQVAVSLYPSRKKVPVKDLEPGINS 1080
QY 1081 POLAKKILSLSEEGSLERHKKQADDTISNASSQLSPSPSOSPRKGYTLABGTYDNF 1140
|
|
|
Db 1081 POLAKKILSLSEEGSLERHKKQADDTISNASSQLSPSPSOSPRKGYTLABGTYDNF 1140
QY 1141 SDSGHSSEISRSIVNSFSFDSVPSLHDERORHSHVSIYETNLGMRMERTMIEPDY 1200
|
|
|
Db 1141 SDSGHSSEISRSIVNSFSFDSVPSLHDERORHSHVSIYETNLGMRMERTMIEPDY 1200
QY 1201 SLGSAAMSGRGLXATATYISSPSTELSDQDGRASLDAASGSGSWTSCSGSHDNT 1260
|
|
|
Db 1201 SLGSAAMSGRGLXATATYISSPSTELSDQDGRASLDAASGSGSWTSCSGSHDNT 1260
QY 1261 OTIOHQSWEFLPFQHTHFYSGDPALMASSHMDQIMFSDSTXKNRNOGRESLEOA 1320
|
|
|
Db 1261 OTIOHQSWEFLPFQHTHFYSGDPALMASSHMDQIMFSDSTXKNRNOGRESLEOA 1320

```

```

QY 1321 QSRASMASTIWGEDSGDTGTTIKRGGKDVSLIEASSSLTSTVTEETKRVPMPAHIAV 1380
|
|
|
Db 1321 QSRASMASTIWGEDSGDTGTTIKRGGKDVSLIEASSSLTSTVTEETKRVPMPAHIAV 1380
QY 1381 ASSTTKGLIARKEGRYRPPTPPGYIGIPITDPDECHSHPARCPDYNALORSRVAR 1440
|
|
|
Db 1381 ASSTTKGLIARKEGRYRPPTPPGYIGIPITDPDECHSHPARCPDYNALORSRVAR 1440
QY 1441 SSDTAGPSSVQOPGHPTSSRPVANKPQMKHKNESDPRLAPYOSOGFSTEDEDEQVSAY 1499
|
|
|
Db 1441 SSDTAGPSSVQOPGHPTSSRPVANKPQMKHKNESDPRLAPYOSOGFSTEDEDEQVSAY 1499

RESULT 2
US-09-911-826A-7
; Sequence 7, Application US/09911826A
; Patent No. US20020143164A1
; GENERAL INFORMATION:
; APPLICANT: Rotin, Daniela and Pham, Nam
; TITLE OF INVENTION: RAS Activator Nucleic Acid Molecules, Polypeptides and
; FILE REFERENCE: DDW-5001-US
; CURRENT APPLICATION NUMBER: US/09/911,826A
; PRIOR FILING DATE: 2002-02-26
; PRIOR APPLICATION NUMBER: PCT/CA00/00042
; PRIOR FILING DATE: 2000-01-20
; PRIOR APPLICATION NUMBER: 2,259,830
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-911-826A-7

Query Match 10.1%; Score 777.5; DB 10; Length 307;
Best Local Similarity 51.3%; Pred. No. 8.2e-40;
Matches 157; Conservative 59; Mismatches 77; Indels 13; Gaps 5;

QY 709 SOISLQSLSTVEVATQLSMNFELFRNIEPTETIDOLFRLKSTSCANLKRFEVINOEF 768
|
|
|
Db 1 SNVHFLHNLNAYELAQTLDDFANFRQIESTEYVDELFLRSLKGYVPMLSKFAELVNRMA 60
QY 769 FWVASELFRKTNOIKRKKIKHFKIALHCRECKNFNSMFAIISGLNLAVARLRTTWEL 828
|
|
|
Db 61 FWVASELCAEHNTVRRKKIVKQFIKIALHCRECKNFNSMFAIISGLNLAVARLRTTWEL 120
QY 829 LPNKYEKLFQDLDLPDPSRNMAKRYNVLSQNL-QPPIITLPFVYIKDLTFLEHGNDSK 887
|
|
|
Db 121 LPNKYQKLFQDLDLPDPSRNMSKYIOLVSAELLAOHPIITPPYVYIKDLTFLEHGNDR 180
QY 888 VDLGVNFEKLRMLAKIRHVGMAVNMND-PAIMFRTRKKKWRSLGSLSGSTNA-- 941
|
|
|
Db 181 VDLGVNFEKLRMLAKIRHVGMAVNMND-PAIMFRTRKKKWRSLGSLSGSTNA-- 240
QY 942 -TYLDAVQIGH---KKVRRSSEFLNAKKLYEDAQARKVQYLSNLELMEDESLQTL 997
|
|
|
Db 241 GTV--IAANAGATIRKRRKSTAAPNPKMFEBQAVRRKAVYLSLKLSDLEDLLHKFS 298
QY 998 LOCEPA 1003
|
|
|
Db 299 LCECPA 304

RESULT 3
US-09-911-826A-4
; Sequence 4, Application US/09911826A
; Patent No. US20020143164A1
; GENERAL INFORMATION:
; APPLICANT: Rotin, Daniela and Pham, Nam
; TITLE OF INVENTION: RAS Activator Nucleic Acid Molecules, Polypeptides and
; TITLE OF INVENTION: Methods of Use

```



Db 1 LLLDPEELAEQLLLDLELFRKIEPSSELLGSVWGKRSKSSPLAPQNLAEATIERENEV 60  
Qy 768 TFWVASLIRET--NQLRMKIKHFIKIALHCRCKNPNMFATISGLNLAPARLRTT 825  
Db 61 SNWVATELKKQTTLKPKRAEVLSEFIVAHCHRELNNFNSLMAIVSALSSPSISRLKKT 120  
Qy 826 WEKLPNKYEKLEFQDLQDLFDS-RNMAYRNVL-----NSQNLQPIIPLFVYIKDLTFL 880  
Db 121 WEKLSKRYKKLFEELLEDLDPSEKRNFKYREALKSCNKSPPVQPCVDFGLYIKDLTFL 180  
Qy 881 HEGN-----DSKVDGLVNEKILRMATAKLRHVGRAVSVMPPALMFRTRKKKWSLSLSQ 936  
Db 181 DEGNDFLENGTGLVNEKRRKIAKILREI-----RQLQSACQ 219  
Qy 937 GSTNATVDAVQGTGHHKRRVRSSEFLNAKKLYEDAMARKVQYLSNLEMEDEESIQTL 996  
Db 220 P-----YNLKPKNRNDIQELLRASPLLEVLEPEEDE--LYEL 253  
Qy 997 SLOCEP 1002  
Db 254 SLRIEP 259

RESULT 6  
US-09-940-836A-6  
Sequence 6, Application US/09940836A  
Patent No. US20020146800A1  
GENERAL INFORMATION:  
APPLICANT: Millennium Pharmaceuticals, Inc.  
APPLICANT: Curtis, Rory  
TITLE OF INVENTION: 48921, A NOVEL HUMAN GTP RELEASING  
FILE REFERENCE: 38155-20031.00  
CURRENT APPLICATION NUMBER: US/09/940,836A  
CURRENT FILING DATE: 2001-08-27  
PRIOR APPLICATION NUMBER: US 60/228,760  
PRIOR FILING DATE: 2000-08-30  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6  
LENGTH: 246  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Consensus amino acid sequence  
US-09-940-836A-6

Query Match 6.1%; Score 474.5; DB 10; Length 246;  
Best Local Similarity 45.8%; Pred. No. 1.2e-21;  
Matches 97; Conservative 39; Mismatches 65; Indels 11; Gaps 3;  
Qy 707 RESQISLLQSLVYEVATQLSMRNFELFRNIEPEYIDDLFKLSKTS---CANLKRFE 762  
Db 4 RKVSLDILHDLPEELAEQLLLDLELFRKIEPSECLGYVWSNREKKGNSPLNLEKFTQ 63  
Qy 763 VINOETFWVASLIRETNQLRMKIKHFIKIALHCRCKNPNMFATISGLNLAPARLRTT 822  
Db 64 RENNISYVATELSEKQQAQAKIIEKIVAOCHRELNNFNSLMAIVSALSSPSISRLKKT 123  
Qy 823 RTTWKLPNKYEKLEFQDLQDLFDSRNMAYRNVLNS-----QNLQPIIPLFVYIKDL 877  
Db 124 KTWKLVVPEKTFEELLEDLDPSEKRNFKYREALKSCNKSPPVQPCVDFGLYIKDLTFL 183  
Qy 878 TFLHEGNDKVD--GLVNEKILRMATAKLRHV 907  
Db 184 TFIHEGNPDYLDNTNLINFEKRMIAKILREI 215

RESULT 7  
US-09-922-199A-9  
Sequence 9, Application US/09922199A  
Publication No. US20020187138A1  
GENERAL INFORMATION:

APPLICANT: Meyers, Rachel  
TITLE OF INVENTION: 15368, A NOVEL HUMAN GTP-RELEASING  
FILE REFERENCE: 381552002500  
CURRENT APPLICATION NUMBER: US/09/922,199A  
CURRENT FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: 60/222,622  
PRIOR FILING DATE: 2000-08-02  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9  
LENGTH: 238  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Consensus amino acid sequence  
US-09-922-199A-9

Query Match 6.1%; Score 468.5; DB 9; Length 238;  
Best Local Similarity 46.8%; Pred. No. 2.6e-21;  
Matches 96; Conservative 37; Mismatches 62; Indels 11; Gaps 3;  
Qy 713 LQSLTVEVATQLSMRNFELFRNIEPEYIDDLFKLSKTS---CANLKRFEVINQET 768  
Db 1 LHLDEPEELAEQLLLDLELFRKIEPSECLGYVWSNREKKGNSPLNLEKFTQRFNNS 60  
Qy 769 FVWASLIRETNQLRMKIKHFIKIALHCRCKNPNMFATISGLNLAPARLRTTWEK 828  
Db 61 YWATELISEKQQAQAKIIEKIVAOCHRELNNFNSLMAIVSALSSPSISRLKKTWEK 120  
Qy 829 LPNKYEKLEFQDLQDLFDSRNMAYRNVLNS-----QNLQPIIPLFVYIKDLTFLHEG 883  
Db 121 VKETIKLFEELLEDLDPSEKRNFKYREALKSCNKSPPVQPCVDFGLYIKDLTFLHEG 180  
Qy 884 NDSKVD--GLVNEKILRMATAKLRHV 907  
Db 181 NPDYLDNTNLINFEKRMIAKILREI 206

RESULT 8  
US-09-911-826A-8  
Sequence 8, Application US/09911826A  
Patent No. US20020143164A1  
GENERAL INFORMATION:  
APPLICANT: Rotin, Daniela and Pham, Nam  
TITLE OF INVENTION: RAS Activator Nucleic Acid Molecules, Polypeptides and  
FILE REFERENCE: DDM-5001-US  
CURRENT APPLICATION NUMBER: US/09/911,826A  
CURRENT FILING DATE: 2002-02-26  
PRIOR APPLICATION NUMBER: PCT/CA00/00042  
PRIOR FILING DATE: 2000-01-20  
PRIOR APPLICATION NUMBER: 2,259,830  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 8  
LENGTH: 270  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-911-826A-8

Query Match 5.2%; Score 399; DB 10; Length 270;  
Best Local Similarity 32.1%; Pred. No. 4.9e-17;  
Matches 98; Conservative 52; Mismatches 95; Indels 60; Gaps 5;  
Qy 716 LSTVEVATQLSMRNFELFRNIEPEYIDDLF---KLSKTSKANLKRFEVINQETFWVA 772  
Db 8 VSARDLAGQLTDHDSLFNSIHQVELHYVLPQHLDVTT-ANLEFRMRFRNELQYWA 66  
Qy 773 SEIRETNQLRMKIKHFIKIALHCRCKNPNMFATISGLNLAPARLRTTWEKLPNK 832  
Db 67 TELGLCPVGPBRAQLLKRFKILAHLEKQKLNLSFPAVMGLSLSAISRLAHLEWRLPKH 126







```
Db 1262 SVSAVQVTRKRDNKSPI--HMSSSSLPSSA-SSAFPRLLKLLDIDPYATQTLVLEH 1318
QY 731 ELERNIEPEYIDDLF--KLRSKTSKANLKREEVINQETFWVASPLLENTOLKRMKII 788
Db 1319 DLRLTTECCLDRAQGTTCNMGSPNITKFLANANTLTNEVSHRTIVQADYKTSKLT 1378
QY 789 KHEIKIALHCRECKNENSMFAIISGLNLAPVARLRTTWKLPKRYEKLPODDLDPDSR 848
Db 1379 QYFVYVAQHCKELNNSSMTAIVSALYSPTIYRLKKTWDLVSTESKDLKLNLMDSKR 1438
QY 849 NNAKYNNVNSONLQPIIPLEFPVIKKDLTFLEHGNDS--KVDGLVNEFKLMIKAEIR 905
Db 1439 NEVYKRELLRSVT-DVACVPEFEGVYLSDLTFTEVGNPDFLHNTNINFSKRRIANIYE 1497
QY 906 HV 907
Db 1498 EI 1499

RESULT 14
US-09-911-826a-10
; Sequence 10, Application US/09911826A
; Patent No. US20020143164A1
; GENERAL INFORMATION:
; APPLICANT: Rotin, Daniela and Pham, Nam
; TITLE OF INVENTION: RAS Activator Nucleic Acid Molecules, Polypeptides and
; FILE REFERENCE: DDW-5001-US
; CURRENT APPLICATION NUMBER: US/09/911,826A
; CURRENT FILING DATE: 2002-02-26
; PRIOR APPLICATION NUMBER: PCT/CA00/00042
; PRIOR FILING DATE: 2000-01-20
; PRIOR APPLICATION NUMBER: 2,259,830
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 10
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-911-826a-10

Query Match 3.4%; Score 260.5; DB 10; Length 249;
Best Local Similarity 29.9%; Pred. No. 1e-08;
Matches 60; Conservative 48; Mismatches 86; Indels 7; Gaps 4;

QY 710 QISLLQLSTVEVATQLSMRNLFERNIEPTDYIDLFLKLSK-TSCANLKREEVINQET 768
Db 2 EITLLTLHPLELARQLTLLEFEMYNKVPSELVGSPTKKDEKSPNLLKIMKHTTNTV 61
QY 769 FWVASPLLENTOLKRMKIIKHEIKIALHCRECKNENSMFAIISGLNLAPVARLRTTWK 828
Db 62 RWTKESTTEANEVEERLAIQRAIEVMVMLELNNGILSIYAAMGTASVYRLRWTFQG 121
QY 829 LPNKRYEKLPODDLDPDSRNNAKYNNVNSONLQPIIPLEFPVIKKDLTFLEHGNDSKY 888
Db 122 LPRRYKRFLECEKELSDD-HLKKYQERLRSIN--PPCVFFFGRYLTNIHLLEGNDLL 177
QY 889 DG--LVNFEKLMIAKEIRHV 907
Db 178 ANTELINFSKRRAVAETIGEI 198

RESULT 15
US-09-911-826a-9
; Sequence 9, Application US/09911826A
; Patent No. US20020143164A1
; GENERAL INFORMATION:
; APPLICANT: Rotin, Daniela and Pham, Nam
; TITLE OF INVENTION: RAS Activator Nucleic Acid Molecules, Polypeptides and
; FILE REFERENCE: DDW-5001-US
; CURRENT APPLICATION NUMBER: US/09/911,826A
```

```
; CURRENT FILING DATE: 2002-02-26
; PRIOR APPLICATION NUMBER: PCT/CA00/00042
; PRIOR FILING DATE: 2000-01-20
; PRIOR APPLICATION NUMBER: 2,259,830
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 9
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-911-826a-9

Query Match 3.3%; Score 254; DB 10; Length 244;
Best Local Similarity 25.9%; Pred. No. 2.4e-08;
Matches 76; Conservative 46; Mismatches 105; Indels 66; Gaps 6;

QY 716 LSTVEVATQLSMRNLFERNIEPTDYI-DLFLRSKTSKANLKREEVINQETFWVASE 774
Db 8 LSAAMELAEQITLLDHTVFRSIPYEEFLGCGMKLDKNERTPYIMKTSQHFNENSNLYASQ 67
QY 775 ILRETNOLKRMKTIKHEIKIALHCRECKNENSMFAIISGLNLAPVARLRTTWKLPNKYE 834
Db 68 INMYADISSRPNAIEKWAVADICRCLHNYNGVLEITSALNRSPIYRLKKTMAKVSQTK 127
QY 835 KLFQDLODLDPDSRNNAKYNNVNSONLQPIIPLEFPVIKKDLTFLEHGNDS-KVDGLVN 893
Db 128 ALMDKLOKTYSSSGRRKNEETL--KNCNPAPVAYLGMVLTDLAFLEEGPNFTTEGLVN 185
QY 894 FEKLRLMIA--KEIRHYGRMA-SVNMDPALMFRTRKKKWSLSLSGSGTNATVLDVAQT 949
Db 186 FSKMRMISHIRIRIQEQYARIDQOP----- 213
QY 950 GGHKKRVRRSSFNLNAKKLYEDAQMARVXOYLSNLEIEMDESLSQTLISLQCEP 1002
Db 214 -----KVIOYLLDKALVIDEDSLYELSLKTEP 240
```

Search completed: February 18, 2003, 21:40:36  
Job time : 303 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2003, 21:22:09 : Search time 40 seconds  
(without alignments)  
1102.624 Million cell updates/sec

Title: US-09-911-826a-2

Perfect score: 7721

Sequence: 1 MKPLAIPANHGVGQGEKHS.....PYSGQSTFEDDEDEQVSAY 1499

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Issued\_Patents\_AA:\*
- 2: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/6CTUS.COMB.pep:\*
- 7: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	354.5	4.6	1596	3	US-09-356-952-3
2	323	4.2	1572	2	US-08-290-731C-5
3	321.5	4.2	1048	2	US-09-356-952-5
4	307.5	4.0	1319	2	US-08-290-731C-2
5	307.5	4.0	1336	2	US-08-290-731C-6
6	303.5	3.9	1333	3	US-09-356-952-2
7	288	3.7	1244	3	US-09-356-952-7
8	285	3.7	418	2	US-08-290-731C-13
9	282.5	3.7	1589	3	US-09-356-952-4
10	277	3.6	911	3	US-09-356-952-6
11	276.5	3.6	666	1	US-08-318-831-3
12	276.5	3.6	814	1	US-08-318-831-2
13	274	3.5	489	1	US-08-318-831-4
14	273	3.5	402	2	US-08-290-731C-14
15	269	3.5	430	2	US-08-290-731C-9
16	268	3.5	464	2	US-08-290-731C-12
17	265	3.4	326	1	US-08-318-831-6
18	264	3.4	357	4	US-09-356-952-1
19	248	3.2	652	1	US-08-318-831-8
20	245.5	3.2	1297	2	US-08-290-731C-4
21	238.5	3.1	768	5	US-08-408-519-2
22	238.5	3.1	768	5	US-08-408-519-2
23	238.5	3.1	852	2	US-08-408-519-5
24	238.5	3.1	852	2	US-08-408-519-5
25	227	2.9	423	5	US-08-290-731C-10
26	201.5	2.6	2101	1	US-08-195-487-4
27	201.5	2.6	2101	5	PCT-US93-06160-4

28	199.5	2.6	423	2	US-08-290-731C-11	Sequence 11, Appl
29	197.5	2.6	2101	1	US-08-466-390-4	Sequence 4, Appl
30	197.5	2.6	2101	1	US-08-470-950-4	Sequence 4, Appl
31	197.5	2.6	2101	1	US-08-467-781-4	Sequence 4, Appl
32	197.5	2.6	2101	2	US-08-483-924-4	Sequence 4, Appl
33	197.5	2.6	2101	4	US-09-452-294-1	Sequence 1, Appl
34	197.5	2.6	3248	5	US-08-353-700-1	Sequence 1, Appl
35	197.5	2.6	3248	5	PCT-US95-16216-1	Sequence 1, Appl
36	197	2.6	1066	4	US-09-541-782-8	Sequence 8, Appl
37	197	2.6	1066	4	US-09-723-820-8	Sequence 8, Appl
38	194.5	2.5	2482	1	US-08-328-254-6	Sequence 6, Appl
39	191.5	2.5	1184	4	US-09-541-782-2	Sequence 2, Appl
40	191.5	2.5	1184	4	US-09-723-820-2	Sequence 2, Appl
41	186.5	2.4	897	1	US-08-095-737-4	Sequence 4, Appl
42	186.5	2.4	897	1	US-08-480-145-4	Sequence 4, Appl
43	186.5	2.4	897	2	US-08-477-389-4	Sequence 4, Appl
44	186	2.4	1164	4	US-08-923-992A-10	Sequence 10, Appl
45	185.5	2.4	10182	4	US-09-134-001C-3159	Sequence 3159, Ap

ALIGNMENTS

```
RESULT 1
US-09-356-952-3
; Sequence 3, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
; APPLICANT: Boriack-Sjodin, Ann
; APPLICANT: Margarit, S. M.
; APPLICANT: Bor-Sogli, Dafna
; APPLICANT: Cole, Philip
; APPLICANT: Kariyan, John
; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
; FILE REFERENCE: 600-1-228N
; CURRENT APPLICATION NUMBER: US/09/356,952
; EARLIER FILING DATE: 1999-07-19
; EARLIER APPLICATION NUMBER: 60/093,631
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1596
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-356-952-3
Query Match
Best Local Similarity 17.1%; Pred. No. 3e-17; Length 1596;
Matches 309; Conservative 204; Mismatches 474; Indels 823; Gaps 64;
QY 12 VMGQEKHSPLADFT-----KLHLTDSLHPQVTHVSSH-----SGCSITSD 53
DB 225 MUNGSEHLPLSPSLPAQRASATYEETVKELHDEKQYORDLMIIRVRELVIIVSD 284
QY 54 SGS-----SLSLDIYATESAGDMD-----LSGLPEPAVSDDEDDIEDIE 95
DB 285 PRELEPIFSNMIMYEVTVTLGSLDVIEMSOEGAPCVGSCFEELAEFEEDVKKYA 344
QY 96 RADPPLMSRPIVDCLEK-----113
DB 345 YDVTQSARDLNNLSKPCGASSLTAGHGFRDAVKYKLLPLLVPCIAFYFDYIKHL 404
QY 114 DPIDRTDDDI---EQLEEFHQPLPAFANMTS-----VRRELCAVMAVAYE 157
DB 405 KDLSSSDDELESEFQVGLLPHLCDEKYMASLSKERQYVSGRRKRL-----AIE 457
QY 158 RAGTIYV-----NDGEELDSWSYILNG-----SVEVTYPDGKAEILMGNS 198
DB 458 RTRELQKVEHWDKDVQGNCFEIRSDLSKSGSKRIWSEKRVFLDGLMWLVCKANT 516
QY 199 FGVSPI-----MDKEYMKGV-----MRTKYVDCCO-----222
```

```

Db 517 KKQPSAGATADYRLKEXFMRVNDINDRPDSDDLKNSFELAPRMQPIVLTKAQNQH 576
QY 223 -----FVCIADODYRILNOVEKNMQ-----VEEGELVAKKEHREL 260
Db 577 HDWADMLAVITXSMRLNHDLSIIDIERKHPLMRPSPEIYKFAVPDSCGNIVLEERES- 635
QY 261 DRGTGRKHIVIGTS-----ERLTMHVEHSVVDPTFIEDPLTYRTFLSSPMVEGKK 315
Db 636 -----AGVPKIKATLCKLIERLTHI-----YADPTEKTLTYRYF-CSFOQLLOL 683
QY 316 LLEMFN--DPSL-----RDKATRV 332
Db 684 LVERFNIPDPSLVQODGTAGAGGCGVGDKEHKNSHREDMKRYKREYQVPQVFLNV 743
QY 333 VLLVNNHFNDEGDPAMTRPLEFENNLEREKMGHRLNLNTACAKAKRRLMTLTPS 392
Db 744 LRHVVDHFFYDFEKDPMI--LEKLNFLF--HVN-----GKSMRK----- 779
QY 393 REAPLPILLGSEKGFIFVDSVDSGSKATEAGLKRGDQILEVNGQNFRIQLSKAMEI 452
Db 780 -----WDSY----- 784
QY 453 LRNNHLSITYKTNLFVFKELLTRLSEKRGAPHLPIGDIKKASXSIPLDAVDVEQV 512
Db 785 -----LKI----- 787
QY 513 IGLEKVKKSKANTVGRNKKLILDKTRISILPQKPYNDIGIGQSDDSIVGLRQTKHI 572
Db 788 --VORKNQEKSN-----KKTIV-----YATG-----HD 808
QY 573 PTLAPVSGTSSNPDLQSHHRLDEFSATPDPQVLRVFKADQOSRYIMISKDTAKE 632
Db 809 PP--PIEHLSVPR----- 820
QY 633 VVIOAIRFPAVTATPDQYSLCEVSVTEBGVIKQRRLPDQLSKLADRIQLSGRYLKNME 692
Db 821 ----- 820
QY 693 TETLCSDEDAQELLRESOISLLOSTJEVATQLSMRNPELEFRNIEPTREYIDDLFKLSK- 751
Db 821 -----DEITLTLHPLELARQTLLEEMKKNVPSLESGPMTKDKE 864
QY 752 TSCANLKRFEVINOETFWVASEILRETNOLKRMKIIKHFIKIALHGRECKNFMSMAFI 811
Db 865 VKSPNLLKIMKHTNVTNRWIEKSTTEAENYEERLAIMQRALEVAMVMELENNFGILSI 924
QY 812 SGLNLAPVALRTWTEKLPNKYEKLFQDLQDLEPDSRNMAKYRVNLSQNLQPIPLP 871
Db 925 AAMGTASYRLRMTFOGLPERYRKFLCECRELSDD--HLKKYQERLSIN--PPCVDFEG 980
QY 872 VIKKDLFLHEGNDKSYDG--LVNEFKLRMAKELRIHVGRNASVNMPPALMFRTRKKMR 929
Db 981 RYLTNLHLEEGNDLIANTELINFESKRKYAELI----- 1015
QY 930 SLGSLSGSTNATVLDVAQOTGHHKRVRRSSFLNAKKLYE-----DAQMARKVQY 980
Db 1016 --GEIQOYQONPYCLNEST-----IROFEQDLDPENGLSDQMS--DY 1055
QY 981 LSNLELDESDLSOTLSQCPATNTLPKNPGDK-----KPYKSEISPAAPRAGS 1030
Db 1056 LYNESLRLEPRGCTVYKPPRKMWHIPKSPGKIPRQONQTNSSSKLSNSTSVAAAAA 1115
QY 1031 QOKQOSLPQOQPPRAHKINQGLQVAVSLYPSRKVKPYKDLPPFGINSQALKITLSL 1090
Db 1116 SSTSTST-----ATASAPSLHASSIMDAP----- 1139
QY 1091 SEEGSLERHKKQADDTISNASSOLSSPTSPQSSPRKGYTLAP-----SGTVDN 1139
Db 1140 -----TAAANAGSGTLAGEQSQHNPFAVSFAVPILEBRNTSSMSGTPQH 1166
QY 1140 F-SNDSGHSIS-----SNRSIVSNSSFDSV-----PVSLLDER 1171

```

```

Db 1187 TRTDQNNGEVSPAPHLPKKRGAVHANNNSTLASASAMDVFPSPALPEHLPPQSLPDSN 1246
QY 1172 RQHSVSIYETNIGMGEMERTMIEPDQYSLGSYAPMSBGRGLXATATVSSPTEELSQ 1231
Db 1247 PFASTDTEAPPSPPL-----PKLVVSP-RHETGNRSP--HGMQNSPHTSTASTVTLT- 1295
QY 1232 DQGDRAISLDAADSGRGSWTSCSSGSHDNIQTIQHRSWETLPPG-----HTHPDYSQD 1284
Db 1296 -----GMSTSGGEELCAGGFYFN--SAHQQPGAVPLSPHIVNPMATNMETRAV 1342
QY 1285 PAGL-----WASSHMDQIMFS-----DHST--KYNQONOS 1313
Db 1343 PPPLPRRKERTESCADMAOKROAPDAPLTPRDELSPPPIPRLNHSTGISYLRQSHQ 1402
QY 1314 RESLEQOSRASMASSTGVYGEDS--EGDGTIKRGGKXVSTEASS-----S 1360
Db 1403 K-----SKEFGNSSLLPNSSIMIRNSAIEKRAATSOPONAAAGPTS 1448
QY 1361 LTVYTEET-----KPYMPAHIAVASTTKGLIARKRGYREPPPT-----PGYIGIP 1410
Db 1449 TTLVTSQAVATDEPLPLPISPAASSTTS-----PLTPAMSPMNPISPH 1496
QY 1411 ITDPFEGSHPARKPPDYVALORSKRVANASDTPAGSSVQOPHGHPITSRPVKNQOWHK 1470
Db 1497 VESTSSSYAH-----QLRMROOQOQOTHPAIVSOHHHATLTP-HHPHONH 1542
QY 1471 PNESDPRLAP 1480
Db 1543 SNPTQSRSP 1552

```

## RESULT 2

US-08-290-731C-5.

; Sequence 5, Application US/08290731C

; Patent No. 5843646

; GENERAL INFORMATION:

; APPLICANT: BOWTELL, David Douglas Lawrence

; TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE

; TITLE OF INVENTION: SON OF SEVENTEEN (SOS) GENE,

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SUGHRUE, MIOT, ZINN, MACPEAK &amp; SEAS

; STREET: 2100 PENNSYLVANIA AVENUE, N.W.

; CITY: WASHINGTON

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/290,731C

; FILING DATE: 17-OCT-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/AU93/00068

; FILING DATE: 17-FEB-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PL0921/92

; FILING DATE: 17-FEB-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: KIT, Gordon

; REGISTRATION NUMBER: 30,764

; REFERENCE/DOCKET NUMBER: Q-36066

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 293-7060

; TELEFAX: (202) 293-7860

; TELEX: 6491103

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

```
; LENGTH: 1572 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-290-731C-5

Query Match      4.2%; Score 323; DB 2; Length 1572;
Best Local Similarity 16.9%; Pred. No. 6.4e-15;
Matches 292; Conservative 196; Mismatches 446; Indels 794; Gaps 61;

QY 12 VMQOQEKHSIPADFT-----KHLTDSLHPQVTHVSSH-----SGCSTSD 53
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 225 MLNQSEAHILPSPLSPAQASATYEETVELHDEKQYRDLMHITRFRRELVLKIVSD 284
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 54 SGS-----SSLSDIYQAFSEAGDMD-----LSGLPETAVIDSEDDDEEDIE 95
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 285 PRELEPFFSINMDIETVTLVLLSLEDVIMSDQSAFCVSGCFEELAEEDFYKKYA 344
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 96 RASDPLMSRDIVRDCEK-----113
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 345 YDVTQSASRDALNNLLSKPGASSLTAGHGFRDAVKYTLKLLVPIGAFVYFDYIKHL 404
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 114 DPLDRIDDDI---EQLEFHHQLPAPANNMTS-----VRRELCAMVFAVVE 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 405 KDLSSQDDDIESTFEYOGLHPLHCDLEKVMASLSKERQVPVSGRVKROL-----AIE 457
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 158 RAGTIVL-----NDGEELDSWSVLNG-----SVEVYTPDCKAEILCMGNS 198
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 458 RTLELOAKVYHMEKDKYQNCNEPIREDLSLSKSGKRIWSEKRVLELFDG-LMVLCAANT 516
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 199 FGVSPT-----MDKEYMKGV--MRTKYDDCO-----222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 517 KQRTPSGATAYDYRLKEKYFMRVINDIRPSDDLKNSFELAPRMQPIVLAKNAQHK 576
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 223 -----FVCIQAODQYRIILNQVEKNQK-----VEEGEITVWKEHREL 260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 577 HDWMADLLAVITKSMRLHDSILODIERKPLRMPSEIYKFAVPDSCGNIVLEERES- 635
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 261 DRTGRKGHIVIGKTS-----ERLTMHVLEHSVDPFTIEDFLTYRTFLSSPMERKK 315
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 636 -----AGVMTIGATICKLIERLTHI-----YADPTFVITLTYTRYF-CSPOQLQL 683
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 316 LLEMFN--DESL-----RDKVTRV 332
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 684 LVERFNIPDSLVYYQDTAGAGMGVGDDKEHKNSHREDMKRYREKYVQVQFRFLNV 743
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 333 VLLAVNNHNFDEGDDPAMTFFLEFFENNEREKMGHLLRLNATCAKAKRRLMTLTKPS 392
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 744 LRHWVDHNFYDFEKDPMI--LEKLLNLE-----HVN-----GKSMRK----- 779
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 393 REAPLPIILGSEKGCIFVDSDVSGSKATFAGLKRGDQILLEVNGNPFENIQLSKAMEI 452
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 780 -----WVDSV-----784
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 453 LRNNTHLSITVKTNLVFKELLTRLSEKRNKGAHLPKIGDIKKASHYIPDLAVDVEQV 512
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 785 -----LKI-----787
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 513 IGEKYNKKSKANTAVGGRNKLKLTIDTRISILPQKPYNDIGIGOSODDSIVGLRQTKHI 572
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 788 --VQRKNEQEKSN-----KKIV-----YAVG-----HD 808
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 573 PTLAPVSGTSSSNPDLQSHHRLDPSATRDLPDYLVRFKADQOSRYIMISKDTAKE 632
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 809 PP--PIHHHSVFN-----820
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 633 VVIOAIRFAVTAATPDQYSLCEVSVTPREGVYIKORRLPDQLSKLADRIOLSGRYLKNME 692
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 821 -----820
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 693 TETILCSDEDAQELLRESQISLQSLSTVEVATQLSMRNFELFRNIETPTETIDDLFKLSK- 751
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 821 -----DEITLTLHLPLELAROLTLLEFEMYNKVKPSSELVGSPTKKDKE 864
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 752 TSCANLKRFEVINOETFWVASEILRETNOJKRMKIIKHFIKIALHCRCEKNFNSFEAT 811
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 865 VKSPNLLKIMKHTNTVTRWIERKSITEAENEEERLAIMQRAIEVMAMVLENNNGILSTIV 924
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 812 SGINLAPVARLRTTWELKLPNKYEKLFODLQDLFDPSSNNMAKYRNVNLSQMLOPPIPLFP 871
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 925 AAMGTASVYRLRWTFQGLPERYRKFLRECRRLSD--HLKQYERLRSIN--PPCVDFFG 980
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 872 VIKKDLTFLHEGNDISKVDG--LVNFEKRLIAKEIRIVGMASVYNNDPALMFTRRKKWR 929
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 981 RYLTNLIHLEEGNPDDLANTLINFSKRRKVAEII-----1015
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 930 SLGSLSGSTNATVLDVAOTGCHKKRVRRSSFLNAKKLYE-----DAQARKVQY 980
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1016 --GEIQOYQNPYCLINEST-----IROFFQDLDPFNGLSKQMS-----DY 1055
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 981 LSNLEMEDEESLQTLSECEPATNPLPKNPQDK-----KPKSETSPVAPRAGS 1030
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1056 LYNESLRIEPRGCKTYKPFPRKMPHILPKSPGIKPRQONTNSSSKLSNSTSSVAAAAA 1115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1031 QOKAQSLPQOQPPRAHKINQGLQVPVAVSLYPSRKVPYKDLPPGINSPOALKITLSL 1090
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1116 SSTATSI-----ATASAPSLHASSIMDAP-----1139
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1091 SEEGSLERHKKQADRTISNASQSLSPTSPQSPRKGTYLAP-----SGTVDN 1139
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1140 -----TAAANAGSGTLAGEQSPQHNPAFVSFAVYIIPERTSSKSGTPQH 1186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1140 F-SDSGHSEIS-----SRSSIVSNSSPDSV-----PVSILDER 1171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1187 TRTDONNGEVSVAAPHLPKPKGAHVMAANNSTLASASAMDVFSPALPEHLPOQSLPSDN 1246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1172 RQHNVSIVETNIGMGEMERTMIEPDQYLSGYAPMSEGRGLATATVYSSPSTELSQ 1231
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1247 PRASDTEAPSPPL-----PKLYVSP--RHETGNKSPF--HGRQMSPTSTASTVYLT- 1295
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1232 DQGRASLDAADSGRGSTSCSSGSHDNIQTIQHRSMETLPG-----HTHFQVSGD 1284
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1296 -----GMSTSGGEECAGGFTYN--SAHQGPQAVPISPPIVNVPMATMETRAV 1342
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1285 PAGIMASSSHMDQIMFSDHSTKYRNQNSRESLEQASRASWASSTGYWGEDSEGDGTI 1344
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1343 PPLP-----PPRKRERTESCADMAQKQQA-----PDAPTL 1372
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1345 KRKGCKVSTIEASSSLTSVTTEETKVPMPAHIAVASTTKGLIAREKRYREPPPTPP 1404
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1373 PPRDG-----ELSPRPPIPRPL--NHSTGISYLRQSHGSKKE-----1406
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1405 GYIGIPITDEPEGSHPARKPPDYNNVALORSRMVARSDD--TAGPSS 1449
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1407 -FVGNSSLLPNTSSIMIRR--NSAIEK-RAAAVSQPNQAAAGPIS 1448
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
US-09-356-952-5
; Sequence 5, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
; APPLICANT: Borlack-Sjodin, Ann
; APPLICANT: Margalit, S. M.
; APPLICANT: Bor-Sogli, Dafna
; APPLICANT: Cole, Philip
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
; FILE REFERENCE: 600-1-228N
; CURRENT APPLICATION NUMBER: US/09/356,952
; EARLIER APPLICATION NUMBER: 60/093,631
; EARLIER FILING DATE: 1998-07-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
```





Qy	344	FEQRPAMTRLEEEENNLEEKKGHN-RLN-----IACAARAKRMLMTLRPSEA	395
Db	354	FE---LKOLEESENDEDECKMKAQTALLNQSMEKISLSLAKRL-----	400
Qy	396	PLPILLGSGSEKGGIPIVDSDSGSKATEGLARGOILEVNGONFENTOLSKAM-ETLR	454
Db	401	-----SSSACRFYSQOM-----KGQOLAIKKNMEI-QKNIDMEGKDIGQCCNEFM	446
Qy	455	NMHLSTIVK--TWLVYFKELLRLRLEEKNGCAPRL-----KIDDI	494
Db	447	EGTLTRVGAKHEHRIFLFDGLM-ICCKSNHGQPRPLPGSASAEYRLKEKFEYRKVOINDK	504
Qy	495	KKASRY--SIDPLADVDEOYI-----GLEVNKKSAAVTGVGNKKTKITIDTRISILPQ	547
Db	505	DDTSEYKHAFPIILKDGNSYTFSAKSAEENMMALLISLOYRSTILERMLD-----	555
Qy	548	KPYNDIGIGOSODDSIVGLRQTHIPLALPVSGTSSNPDLLOSHNRIL-DESATPDLP	606
Db	556	-----VYVLOEKEEEOGRPLPSAEVYRFAPPD--SEENILFEENYQKAG	597
Qy	607	DOYLR--VEKADQOSYIMISMDTAKEVIOAIREFAVTATPPDOYSLCEVSYRBEVI	663
Db	558	IPITAGVLELLIRLTYHMIADP-----NVRFPFLT--YRSCRPOELLSLI	645
Qy	664	KORRLPOLSKLADRIOLSG-----RYLKNMETETLS-----D	699
Db	646	ERFEIPEPEPTADRIALENGDDPLSAELKRFKKEYIQVOYLRVLNVCNHWHEHNFDE	705
Qy	700	EDAOELRESO-----	710
Db	706	RDADLLQMEFFIGTVGKAKMKWVESITKIQRKKIADNGBNGHNTPOSSPTVEWHI	765
Qy	711	-----ISLQSTVEVATQOLSMRNFELFRNIPEPTYIDLFKRSK-TSCANLKRE	761
Db	766	SRPGHIEFDLLTHPIEIRARQLTLESDELRAVQOSSELVGSWTWKEDKINSPLMKMI	825
Qy	762	EYINQETFWVASIELRETNOIKRMKIIKHFIKALCHCECKNFSMAFIISGLNLAPVAR	821
Db	826	RHTNLTLMFEKCIIVETNEEERAAVYSRIEILLOFOELNENGEVLEVASMNSPYAR	885
Qy	822	LRTMWERKPNKYEKLFODLOLDPD--SRMAKYRVNLSONLOPIIPLFPIKIDLTF	879
Db	866	LDHFEQJPSQOKKILBEAHLELSDHKYTLAKLRSI-----NPPCVFPGIYTLNLIK	939
Qy	880	LHEGDSKVDGLVNFEXLRMLAKEIRHVGMASVYNDPALMFEKTRKKRKSJGSLSGST	939
Db	940	TEEGNP-----EVLRRRKGEL-----INSKRRRAVEITGELIQOYN	976
Qy	940	NATYVDVAQOTGGHKKRRVRSSFLNAKLYEDA-----QARAKYKQYLSLMEDEBSIQ	994
Db	977	OPYCLIREP-----DIKFFENLPMGMSNEKETDILFKRSLDI-----	1016
Qy	995	TLSLOCEBA-TNTPLPKNPGDK--PYKSE-TSPVAPRAGSOOKAQLPQOQOPPAHK	1048
Db	1017	-----EERHHPKPLPRP--KXSYPLAKSGVAPRNPROTGMH--PTPLQOEP-----	1066
Qy	1050	INOGLQVPAVSLYRSKKKVPKYDLPPPGINSPOALKILSLSEEGSLERHKKAOEDTJNS	1109
Db	1061	-----RKISYSRIP-----ESETESTASA	1079
Qy	1110	ASS---OLSSPPTPOSSPRKGYTLASGVNDPFSOGSHSETSSRSIYNSNSPSPVPS	1166
Db	1080	PNSPPTPLTPPPASGTSSNDVCSVPS--DHSASPFHSRSASVSSISLSGTOBEVAP	1133
Qy	1167	LHDERORHVSIVETN--LGMGRMERTMIEPPDOYSLGSAVPMSEGLGATATVVIS	1222
Db	1137	PPVPRRRPRPESAPSSPSKIMSKHLDSPALPBPPOPTMAVSP-----KYSISDRSTISD	1199
Qy	1224	P	1224
Db	1193	P	1193

```

US-08-290-731C-6
; Sequence 6, Application US/08290731C
; Patent No. 5843646
; GENERAL INFORMATION:
; APPLICANT: BOWTELL, David Douglas Lawrence
; TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
; TITLE OF INVENTION: SON OF SEVENLESS (MSOS) GENE,
; TITLE OF INVENTION: AND MSOS POLYPEPTIDES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 PENNSYLVANIA AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290.731C
; FILING DATE: 17-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU93/00068
; FILING DATE: 17-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PL0921/92
; FILING DATE: 17-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KIT, Gordon
; REGISTRATION NUMBER: 30,764
; REFERENCE/DOCKET NUMBER: 0-36066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1336 amino acids
; TYPE: amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-290-731C-6

Query Match 4.0%; Score 307.5; DB 2; Length 1336;
Best Local Similarity 19.6%; Pred. No. 7e-14;
Matches 224; Conservative 169; Mismatches 409; Indels 339; Gaps 47;

236 NOVERKMKVEEGEFT-VWVKEHRE--LDRGTGRGHIVIKGTSRLTMHVLVEHSVDP 292
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
257 NDVENFTSIVIHHEISVYLGHIEDYVMTDEGSPHVLVGSCEFDEL-----ELADP 311
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
293 --TFIEDFLLT--YRTFTLSSPMVEVSKLL-----EMFNDPSLRDVTVRVLLMVNNHEND 343
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
312 YESYADILRPFNGHFLISQLSKPGALYLQSIGGFKE-AVQYVLRLLAPVYHCLHY 370
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
344 FEGDPMRTFLEDFENNLREKMGCHL-RLN-----IACAATAKRLMTLTPRSKEA 395
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
371 FE---LLKOLEKRSDEQDEKCECMQAITALLNVOSGMERICSLSAKRL----- 417
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
396 PLPFIILGSEKGFGLFVNSVDGSKATPEAGLKRGDQILEVNGQNFENITOLSKAM-ELIR 454
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
418 -----SEACRFYSQM-----KGNOLAKKNNET-QKNIIDWEGKDIQCCNEFTM 463
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
455 NNTHTSIYK--TNLFVKELLTRLSEEKRGAPHLP-----KIGDI 494
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
464 EQLTRVVGKHHRRHFTLFPGLM--ICCKSNHGQPRIPGASSAEYRLUKERFMRKVOINDK 521
||| : : : : : : : : : : : : : : : : : : : : : : : : : :

```





```

; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,731C
; FILING DATE: 17-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU93/00068
; FILING DATE: 17-FEB-1993
; PRIOR APPLICATION DATA: P00921/92
; FILING DATE: 17-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KIT, Gordon
; REGISTRATION NUMBER: 30,764
; REFERENCE/DOCKET NUMBER: Q-36066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-290-731C-13

```

```

Query Match          3.7%: Score 285; DB 2; Length 418;
Best Local Similarity 27.2%: Pred. No. 5.2e-13;
Matches 91; Conservative 57; Mismatches 131; Indels 56; Gaps 9;

```

```

QY 680 QLSRRYVILKNNMET-----ETLCSDEDAOELLRESQISLQSLSTVEVATQLSMRNFE 731
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 123 EVNCKFKNGEANTAPKTLIDQICODHYSGLYSTTE-SILAVDPYLFATQTLIEHE 181
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 732 LERNIEPTVEYIDLK--LRSKTSCANLKRFEVINOETFWVASILLRETNQIKRKKII 788
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 182 IYCEITTFDCLQIKMKNYTKSYGASPGLNIEIFANKLTNFIYSYVKEDSKSRKRL 241
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 789 KHEFKIALHGRECKNPNMFIIISGLNLAPVARLRTWEXLPMKYEXKLFODLDPDSR 848
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 242 SHEFTFLAEYCKRNPNFSSMTDIIISALYSSPIYLRLEKTMQAVIPQTRDLQSLNKLMDPK 301
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 849 NMAKYRVNLSQNLQPIILPFLPYIKKDLTFLEHGNDKVDGLVNFELKRLMIKAEIRHVG 908
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 302 NFIVNRNLKSLH-SAPCVPFEGYVLSDLFTDSGNP---DYLVLEHGLKGVHDEKKYIN 357
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 909 RMASVNMDPALMFTRRKKRKSLSLQSGSTNATVLDVAQTGGCHKRRVRSSFLNAAKLY 968
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 358 -----FKKRSR-----LVDTLQ-----EIIYFKKTH 378
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 969 EDAQMARKVQYLSN-LELEMDSESLQTLSDCEP 1002
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 379 YDFTKDRVTIECISNSLENIPHIKQYQSLITEP 413
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

RESULT 9
US-09-356-952-4
; Sequence 4, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
; APPLICANT: Boriack-Sjodin, Ann
; APPLICANT: Margarit, S. M.
; APPLICANT: Bor-Sogli, Dafna
; APPLICANT: Cole, Philip
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE

```

```

; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 600-1-228N
; CURRENT APPLICATION NUMBER: US/09/356,952
; CURRENT FILING DATE: 1999-07-19
; EARLIER APPLICATION NUMBER: 60/093,631
; EARLIER FILING DATE: 1998-07-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1589
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; US-09-356-952-4

```

```

Query Match          3.7%: Score 282.5; DB 3; Length 1589;
Best Local Similarity 22.1%: Pred. No. 6.6e-12;
Matches 133; Conservative 108; Mismatches 244; Indels 117; Gaps 22;

```

```

QY 333 VLLMVNNHFNDEGDPAMTRFL--EEFENNLREKMGHLRLNLTAQ----- 377
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 988 VLVSQDYEFENNL-----FLNNSDDFKDACEK-----YVEISNLACITVDQLIBEREN 1035
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 378 ----AAKAKRRMLTLKPSREAPLPTLLGSGKRGIF-VSDVDSGSKATEGKRGQD 432
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1036 LLNYAARMKNNLTAE-----LLKGEQEKWFDYSEDYSDDDSENDAIIR--DDE 1083
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 433 ILEVNGQNFENIQSKAMELIRNNHTLSITVKTNLVYFKELTRLSEKRNCAHLPKIG 492
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1084 L-----GSEDIYERKANIEKN-----LPWLTSDYELSYLYDSGKRGSTKEA 1128
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 493 DIRKASYSLPDLAVDVEOYIGLEKYNKSKANTVQGRNKKILDKTRISILPQK--PY 550
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1129 LIEHLTSHELVDAFAFNTWMLTFERSI-----LTTREFYALI--YRYNLYPREGLSY 1178
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 551 NDIGIGSQDSDSYLQRTGHIPTALPVSGTSSNPDLQSHNRILDSATPDLDPQYL 610
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1179 DDVNIWIEKSNPKCKVNVIMKRFLOTYTRNRYEPGI---PLLNFAK----- 1225
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 611 RVFKADQSRVIMISKDPTAKEVIAQIRFAVATAPDQYSLCEVSVTEGVYKQRLPD 670
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1226 -----MYSEKIPADDLQKINCKLINENKE-----PYDP---KQD---D 1261
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 671 QLSKLADRIQSGRYVYKNNMETETLCSDEDAOELLRESQISLQSLSTVEVATQLSMRNFE 730
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1262 SVSAVQQTFRDNKSPI--HMSSSLPSSA--SSAFELKRLKLLDIDPYATVATQLVLEH 1318
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 731 ELERNIEPTVEYIDLK--KLRSKTSANLKRFEVINOETFWVASILLRETNQIKRKKII 788
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1319 DLYRITMFECLDAWQTKYCNMGGSFNITKFLANANTLTNFEVSHITVQADVKTTSKLT 1378
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 789 KHEFKIALHGRECKNPNMFIIISGLNLAPVARLRTWEXLPMKYEXKLFODLDPDSR 848
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1379 QYFVTVAQHCKELNPNFSSMAIYSAISPIYRLKKTWDVYESKDILKNLNNLMSKR 1438
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 849 NMAKYRVNLSQNLQPIILPFLPYIKKDLTFLEHGNDKVDGLVNFELKRLMIKAEIR 905
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1439 NEVYKRELLRSVT-DVACVPEFGYVLSDLFTFVGNDFLNHSTNIINFSKRTKIANYIE 1497
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 906 HV 907
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1498 EI 1499
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

RESULT 10
US-09-356-952-6
; Sequence 6, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
; APPLICANT: Boriack-Sjodin, Ann
; APPLICANT: Margarit, S. M.
; APPLICANT: Bor-Sogli, Dafna
; APPLICANT: Cole, Philip
; APPLICANT: Kuriyan, John

```

;; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE  
;; FILE REFERENCE: THEREOF  
;; FILE REFERENCE: 600-1-228N  
;; CURRENT APPLICATION NUMBER: US/09/356,952  
;; CURRENT FILING DATE: 1999-07-19  
;; EARLIER APPLICATION NUMBER: 60/093,631  
;; EARLIER FILING DATE: 1998-07-21  
;; NUMBER OF SEQ ID NOS: 14  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO: 6  
;; LENGTH: 911  
;; TYPE: PRF  
;; ORGANISM: Schizosaccharomyces pombe  
US-09-356-952-6

Query Match 3.6%; Score 277; DB 3; Length 911;  
Best Local Similarity 28.0%; Pred. No. 7e-12;  
Matches 78; Conservative 49; Mismatches 122; Indels 30; Gaps 7;

QY 650 YSLCEVSPREGVYIKORRLPDLSKL-----ADRIQSGRYLKNMTEITLCSDEDA 702  
DB 594 YKPEFVYVPH-FASAEELHSHLHSTRSHKMLEGK--ELSOLEDLSLHNSP 649  
QY 703 QELLRSSQISLQDSTVEVATQLSMRNFELRNIEPTVIDLFLKRSKTSKANLKRFE 762  
DB 650 DPLIKDELVL-LPREIAKQCLIEFQSFHSIRIQFTKIMD-----NLNRFSP 700  
QY 763 VINGEFW-----VASILRETNOIKRMKTIKHFIRKIALHCCECKNNSPFAITSG 814  
DB 701 KEKTSFFYLSNLHVNFTETIVQEEPRRTNVAIFYQCDYLRELNNPASLFSIISAL 760  
QY 815 NLAPVARLRTTWKELPKYKELFQDLODLEDPSPRNMAKYNNVNSONLQPIIPLFVYIK 874  
DB 761 NSSPIHRLKRTWANKNTLASFELNLNLLEAKKNPSNYDCL--ENCVLPVPLGVYF 818  
QY 875 KDLFLHEGNDKVDGLVNFELKLMIAKETIRHVGMAV 913  
DB 819 TDLFLTKGNKDNFQNMINDKRTKVTYRLINEIKKFSQV 857

RESULT 11  
US-08-318-831-3  
; Sequence 3, Application US/08318831  
; Patent No. 5656595  
; GENERAL INFORMATION:  
; APPLICANT: Schweighofer, Fabien  
; APPLICANT: Tocque, Bruno  
; TITLE OF INVENTION: PEPTIDES HAVING A GDP EXCHANGE FACTOR  
; TITLE OF INVENTION: ACTIVITY, NUCLEIC ACID SEQUENCES CODING FOR SAID PEPTIDES,  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhone-Poulenc Rorer Inc.  
; STREET: 500 Aircola Road, 3C43  
; CITY: Colleegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19426  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: Macintosh  
; SOFTWARE: Word 5.1 (Epo PatentIn)  
; CURRENT APPLICATION DATA:  
; FILING DATE: 19 October 1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR92/04827  
; FILING DATE: 21-APR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, Julie K.  
; REGISTRATION NUMBER: P-38,619

;; REFERENCE/DOCKET NUMBER: ST92033-US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (610) 454-3839  
;; TELEFAX: (610) 454-3808  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 666 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
US-08-318-831-3

Query Match 3.6%; Score 276.5; DB 1; Length 666;  
Best Local Similarity 21.0%; Pred. No. 4.7e-12;  
Matches 162; Conservative 114; Mismatches 294; Indels 203; Gaps 28;

QY 277 ERLTMHVEHSVDPTEFIEDFLTYRTFLSSPMEVGKLLLEWNPDSLKDQYTRVLLW 336  
DB 54 ERLT-----DLRFLSDLFLHYSYRVF--TTAIVLDKILITIKKRISAIPARSIELLF 107  
QY 337 VNNHFD-FEGDPAMTRFLEEFENNLEREMGHLRLNLTACAKAKRRLMTLTKSREA 395  
DB 108 ASQNNKLLYGEPP-----KSPRATRFSSPPPLSTIKTSSPSRRKLSLNP--- 155  
QY 396 PLRFILLG-----SEKFGIIFVDSVDSKATFAGLKRQDILEVNGQNFENIQ 445  
DB 156 ----IITGKALDIALSCLNSNGITGYSAMSPFSKAT-----LD 191  
QY 446 LSKAMEILRNHTLSTVTKNLFVFEKELLRLSEERKNAPLPIKIGDI--KKASRYSP 503  
DB 192 TSK-----LVSSSFYTKIPDE-GDTPEKPEPDSALSKOSEVSMR 232  
QY 504 DLAVDVOYIGLEKYVKKSKANVTGGRNK--LKILDKRISILQPKPYNDIGIOSOD 560  
DB 233 EES-DIDQ-----NOSDGDGTETSPYKSPYKSKNNSSEPLFSTYN--GVWVTS 283  
QY 561 DSIVGLRQTKHPTALPVSGTSSNPDLQSHHRIIDFSATPDLDQVLRVFKADQOS- 619  
DB 284 REIDNNRSALSAFAIATAGANEGPNKEKYRRMSLASAG-----PPPOGRNG 333  
QY 620 -RYIMISKDTTAKENVYQAIREFAVATPD---QYSLCEVSPREGVYIKORRLPDQSLK 674  
DB 334 DKPEFVIRRAATNR--VLNVLRHWSKHSODFTNDELCKVIGFLLEVMDPELTQERK 391  
QY 675 LADRIQSGRYLKNMTEITLCSDEDAQELLESQISLQ-----STV 719  
DB 392 AANIT-----IFTLQEDPDG-----DNOITLLEITQMAEGYKAEFENHSAL 433  
QY 720 EVATQLSMRNFELRNIEPTVEYI-DQLFKL-----RSKTSKANLKRFEVINOETFWASE 774  
DB 434 EIMDQLTLDHLVFKKIPIYEEFGQGMKLEKNERPIYIKTKRHFNDISN----LIASE 489  
QY 775 ILRETNOIKRMKTIKHFIRKIALHCCECKNNSPFAITSGNLAPVARLRTTWKELPKY 834  
DB 490 IIRNEDINARVASIEKWAVADICRLHYNVALETSSNRSRAIFPLKTKWILKVSQTK 549  
QY 835 KLFQDLODLEDPSPRNMAKYNNVNSONLQPIIPLFVYIKDLFLHEGNDKSV-DGLVN 893  
DB 550 ALIDKLOKLVSSGREGKNEAL--KNCDPPCVYLGMYLTDLAFLIEGPNYTEGGLVN 607  
QY 894 FEKLIMIAKETIRHVGMAVNMDDPALMFRTRKKKWSLGLSOGSTNATVLDVAQTGCHK 953  
DB 608 FSKMRMISHIREL----- 621  
QY 954 KVRRSSEFLNAKKLYEDAOARVVOYLSLLEMDSESLQTLSCCEPATNT 1006  
DB 622 ----RQFOQTAVKIEHQ----KVTOYLLDOSFVMEDESLYESSLRIEPLPT 666

RESULT 12  
US-08-318-831-2  
; Sequence 2, Application US/08318831  
; Patent No. 5656595  
; GENERAL INFORMATION:



```

Db 26 NKIPDEGDT-----PEKPEDPSAL--SKOSEVSMREESDIDONOSDDGDTETSP----- 74
QY 591 QSHRILDFSATPDLPDQVLRVFKADQOSRYIMSKD-----TTAKEV-----VIAIARE 640
Db 75 -----TKSPPTPKSV-----KKNKSEFPLESTINGVMTSCRELDNNRSALSASA 121
QY 641 FAVT-----ATPDQYSLCEVSTPEGVIKORRLPDP-----LSKL 675
Db 122 FATATGANGCPTNKEKYRRMSLASAGFPDPQNGDKEFVIRRAATNRVLNLRHWYSKH 181
QY 676 A-----DRIQLSGRYLLKNNMETETLCSDP--DAQELR-----ESQISLLDL--- 716
Db 182 SQDEFNDELKCVIGFLEFVMDPELITQERKAANIIRTLQEDPGDNOITLEETQW 241
QY 717 -----STVEVATQLSMRNPELFRNIPTETI--DDLKL-----RSKTSKANLKR 759
Db 242 AEGYKAPFENHSALEADLTLDLVFKKITYEEFFGCGMKLEKNEKTPYIMKTKH 301
QY 760 FEEVINOEFWVASIELRETNOILKRMKIKHFKIALHCECKNFNSMFAIISGLNAPV 819
Db 302 FNDISN-----LIASEIRNEDINARVSAIEKVAADICGLNINAVLEITSMNRSAT 357
QY 820 ARLTTEYEKLPNKYEKLFODLQDFDPSRMATYRVNLSONLOPPIIPLFPYIKDLTF 879
Db 358 FRLLKTLKYSKQTKALIDLQKLVSSEGRFKNLREAL--KNCDPPCVPLGYLTDLAF 415
QY 880 LHHGNDKSV--DGLVNFELKLMIAKEIRHVGRMASVNMDPALMFRTRKKKRSLSGSLSQS 938
Db 416 IEBCTPYTDEGLVNSKMMISHIIREI----- 444
QY 939 TNATVLDVATGCHKVRSSFLNAKLYEDAKOMARKVQYLSNLELMEDESLQTL 988
Db 445 -----KQFOOTAKIRHQA-----KVIQYLLDOSFVWDESLVSSL 481
QY 999 QCEPATNF 1006
Db 482 RIEPKLPT 489

```

RESULT 14  
US-08-290-731C-14  
; Sequence 14, Application US/08290731C  
; Patent No. 5843646  
; GENERAL INFORMATION:  
; APPLICANT: BOWTELL, David Douglas Lawrence  
; TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE  
; TITLE OF INVENTION: SON OF SEVENTLESS (MSOS) GENE,  
; TITLE OF INVENTION: AND MSOS POLYPEPTIDES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
; STREET: 2100 PENNSYLVANIA AVENUE, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/290,731C  
; FILING DATE: 17-OCT-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/AU93/00068  
; FILING DATE: 17-FEB-1993  
; PRIOR APPLICATION DATA: P10921/92  
; APPLICATION NUMBER: P10921/92  
; FILING DATE: 17-FEB-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KIT, Gordon

```

; REGISTRATION NUMBER: 30,764
; REFERENCE/DOCKET NUMBER: 0-36066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-290-731C-14

```

Query Match 3.5%; Score 273; DB 2; Length 402;  
Best Local Similarity 28.9%; Pred. No. 3.8e-12;  
Matches 80; Conservative 49; Mismatches 122; Indels 26; Gaps 8;

```

QY 650 YSICEVSTPEGVIKORRLPDLQSLK-----ADRIQLSGRYLLKNNMETETLCSDEDA 702
Db 102 YKFEVYVYVH--FASAEELHSLSHLHPSRKSHKMLEGK---ELSQLELDLSLHNSP 157
QY 703 QELRESQISLQDSTVEVATQLSMRNPELFRNIPTETI----DDL--FKLSKTSKAN 756
Db 158 DPLIYKDELVL--LPREIAKQCLIEFGQSFHSISRIQFLTKIMDELNRSPEKSTFY 216
QY 757 LKRFEEVINOEFWVASIELRETNOILKRMKIKHFKIALHCECKNFNSMFAIISGLN 816
Db 217 LS-----NHLVNFVETIYQEEPRRTVLAIFYOVCDYLRLELNFASLFSISLNS 270
QY 817 APARLTTEYEKLPNKYEKLFODLQDFDPSRMATYRVNLSONLOPPIIPLFPYIKDL 876
Db 271 SPILRLKRTANLNSKTLASFEELNNLTARKNFSNTRDCL--ENCVLPCVPLGYFTD 328
QY 877 LTFLLHGNDSKVDGLVNFELKLMIAKEIRHVGRMASV 913
Db 329 LTFLLKGTGNKDNFQNMIFDKRTKVTIRLNEIKKFGSV 365

```

RESULT 15  
US-08-290-731C-9  
; Sequence 9, Application US/08290731C  
; Patent No. 5843646  
; GENERAL INFORMATION:  
; APPLICANT: BOWTELL, David Douglas Lawrence  
; TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE  
; TITLE OF INVENTION: SON OF SEVENTLESS (MSOS) GENE,  
; TITLE OF INVENTION: AND MSOS POLYPEPTIDES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
; STREET: 2100 PENNSYLVANIA AVENUE, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/290,731C  
; FILING DATE: 17-OCT-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/AU93/00068  
; FILING DATE: 17-FEB-1993  
; PRIOR APPLICATION DATA: P10921/92  
; APPLICATION NUMBER: P10921/92  
; FILING DATE: 17-FEB-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KIT, Gordon

REGISTRATION NUMBER: 30,764  
REFERENCE/DOCKET NUMBER: Q-36066  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 293-7060  
TELEFAX: (202) 293-7860  
TELEX: 6491103  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 430 amino acids  
TYPE: amino acids  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-290-731C-9

Query Match 3.5%; Score 269; DB 2; Length 430;  
Best Local Similarity 18.3%; Pred. No. 8.4e-12;

Matches 126; Conservative 74; Mismatches 127; Indels 360; Gaps 20;

QY 272 IKGTS-----ERLTMHVAEHSVVDPTFIEDLLTYRTFLSSPMVEGKILEMEN--DPS 324  
DB 1 IKGATLCKLIERLYTHI-----YADPTFVTRFTLYTYRF--CSPOOLQLLVERFNIPDS 54  
QY 325 L-----RDKVTFRVLLVNNHEND 343  
DB 55 LVYQDTGAGAGMGVGCGDKHKSHREDKRYRKEYVQVQFRVLNVLRHWDHHEFD 114  
QY 344 FEGBPAMTRFLEEFENNLEREKMGHLRLNIACAKAKRRLMTLTKESEAPLPFILIG 403  
DB 115 FEKDPML--LEKLNFLFLE-----HVN-----GKSMRK----- 139  
QY 404 GSEKFGIFVDSVDSGSKATEAGLKGDLLEVNGQNFENIQLSKAMEILNNHTLSITV 463  
DB 140 -----WDSV-----LKT-- 147  
QY 464 KTNLFVKELLTRISEKRNAGAPLPIKIGDIKASRYSIPLAVDEQVIGLEKYNKRSK 523  
DB 148 -----VQRKNEQEK 156  
QY 524 ANTVCGRNKKLKKILDKTRISILPKPYNDIGIGSQSDSIYGLKQTKHIPTALPVSCTLS 583  
DB 157 SN-----KKIV-----YAVG-----HDP--PLEHLS 177  
QY 584 SGNPDLLQSHRILDFSATPDLPDQVLRFKADQOSRYIMISKDTTAKENVIAIREFAV 643  
DB 178 VPN----- 180  
QY 644 TATPDQYSLCEVSVTEGVIAQRLPDLQSLADRIQLSGRYLLKNNMETETLCSDEDAQ 703  
DB 181 ----- 180  
QY 704 ELLRESQISLLQSLVVEVATQLSMNFELFRNIEPTFYIDLFLKRSK--TSCANLKRFE 762  
DB 181 -----DEITLLLPLELAPLQTLLEFEKYNKVPSELVGSPPWKDKKEVSPNLKIMK 235  
QY 763 VINOETFWVASEILRETNOQLKRMKIKIHFIKIALHCRCKKNFSMAIISGLNLAPVARL 822  
DB 236 HTTNVTRWTEKSI TEAEVNEERLAIMQRAIEVMWMLNNNGILSIVAAMGTSVYRL 295  
QY 823 RITWEKLPNKYEKLFQDLODLFDPSRNNAKYRNVLNSQNLQPPITLPFVIAKKDLTFLE 882  
DB 296 RMTFOGLPERYRKFLKEGRELSD--HLKKYQERLRSIN--PPCVPFGRYLTNLHLEE 351  
QY 883 GNDSKYVDS--LVNPEKLMIAKEIRHV 907  
DB 352 GNPDLANTELINFSRRKVAEIIIGET 378

Search completed: February 18, 2003, 21:27:58  
Job time : 53 secs



GenCore version 5.1.13  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2003, 17:34:25 ; Search time 86 Seconds  
(without alignments)  
2322.589 Million cell updates/sec

Title: US-09-911-826A-2

Perfect score: 7721  
Sequence: 1 MKPLAIPANHGWMGQGEKHS.....PYGSQGFSTEDDEQVSAY 1499

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database : A.Geneseq.101002:\*

- 1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7721	100.0	1499	21	AAB07792
2	4645	60.2	1675	21	AAAB42658
3	4644.5	60.2	1651	23	ABG66725
4	2352.5	30.5	1551	22	ABR64459
5	1508	19.5	286	20	AAAB8983
6	1508	19.5	286	22	ABR51059
7	676	8.8	884	21	AAV70964
8	671	8.7	881	21	AAV70965
9	664.5	8.6	849	21	AAV70968
10	637	8.3	957	22	ABR59888

11	545.5	7.1	262	21	AAB07793
12	543	7.0	119	22	AAAB5934
13	543	7.0	119	22	AAAB72102
14	543	7.0	119	22	AAAB32364
15	543	7.0	119	23	ABG41916
16	538.5	7.0	152	21	ABAB43070
17	538	7.0	519	22	ABG09372
18	475.5	6.2	261	22	AAE13102
19	475.5	6.2	261	23	AAU99911
20	474.5	6.1	246	23	AAE22105
21	431	5.6	456	22	AAAB78706
22	431	5.6	456	22	AAAB39061
23	418	5.4	291	21	AAV70967
24	409	5.3	338	21	AAV99651
25	354.5	4.6	1596	21	AAV68821
26	321.5	4.2	1048	21	AAV68823
27	321	4.2	556	21	AAV70966
28	312.5	4.0	167	22	AAAB01890
29	310.5	4.0	1336	16	AAAB84638
30	309	4.0	1262	23	ABR09768
31	308.5	4.0	227	23	AAU99910
32	307.5	4.0	1319	14	AAAB47043
33	303.5	3.9	1333	21	AAAB19623
34	303.5	3.9	1333	21	AAV68820
35	299.5	3.9	612	22	ABR71554
36	298	3.9	57	20	AAAB8986
37	298	3.9	57	22	ABR51062
38	296.5	3.8	472	22	AAAB67823
39	294	3.8	1077	16	AAAB91597
40	291	3.8	473	23	ABR04984
41	290	3.8	1275	23	ABR09767
42	288	3.7	1244	21	AAV68825
43	288	3.7	1244	23	ABR09769
44	287.5	3.7	472	23	AAE22102
45	285	3.7	466	22	AAAB68566

#### ALIGNMENTS

RESULT 1	AAAB07792	standard; Protein; 1499 AA.
ID	AAAB07792	
AC	AAAB07792	
DT	07-NOV-2000	(First entry)
DE	A murine guanine nucleotide releasing factor 4.	
XX	guanine nucleotide releasing factor 4; GRF-4; Ras activator;	
KW	Nedd4 ubiquitination; cell metabolism; cell proliferation; cancer;	
KW	cell differentiation; cell transformation; neuronal disorder.	
XX		
OS	Mus sp.	
XX		
FH	Key	Location/Qualifiers
FT	Domain	135..253
FT	Domain	/note="CNP-binding domain"
FT	Domain	266..322
FT	Domain	/note="REM domain"
FT	Domain	386..470
FT	Domain	/note="PDZ domain"
FT	Domain	594..692
FT	Domain	/note="RA domain"
FT	Domain	712..899
FT	Region	/note="CDC25 domain"
FT	Region	1403..1406
FT	Region	/note="py motif"
FT	Region	1425..1428
FT	Region	/note="py motif"
PN	WO200043510-A2.	

Partial murine qua  
Human brain expres  
Human bone marrow  
Peptide #6401 enco  
Human peptide enco  
Human ORF ORF2834  
Novel human diago  
Human phospholip  
Human guanine nucl  
Human 48921 guanin  
Human protein seq  
Human polypeptide  
Rat Ras signalling  
Human GTPase assoc  
Amino acid sequenc  
Amino acid sequenc  
Human CAMP-GPFI fr  
Human polypeptide  
MSO1 protein. Mu  
Amino acid sequenc  
Human 47476 consen  
Mammalian son of s  
Human son of seven  
Amino acid sequenc  
Drosophila melanog  
Polypeptide fragme  
Human secreted pro  
Human guanine-nucl  
C3G protein. Homo  
Human new ras gene  
Amino acid sequenc  
Amino acid sequenc  
Amino acid sequenc  
Human GTP releasin  
Human GTP-binding

XX 27-JUL-2000.  
PD  
XX 20-JAN-2000; 2000MO-CA00042.  
PF  
XX 20-JAN-1999; 99CA-2259830.  
PR  
XX (HSCR-) HSC RES & DEV LP.  
PA  
XX  
PI Rolin D, Pham N;  
XX  
XX WPI; 2000-499228/44.  
DR N-PSDB; AAA59383.  
PT  
XX Nucleic acids encoding guanine nucleotide releasing factor-4 useful for  
the treatment of cancers and neuronal disorders -  
PS  
XX Claim 9; Fig 19a; 89pp; English.  
XX  
XX The present sequence represents a murine guanine nucleotide releasing  
factor (GRF)-4 (Ras activator) polypeptide. GRF4 activates Ras both  
in vitro and in vivo. It directly binds cyclic adenosine monophosphate  
(cAMP) directly via its cAMP-BD (cAMP/guanine monophosphate (cGMP)  
binding domain). GRF4 directly connects cAMP-generating (e.g. G protein  
coupled receptors) or cGMP-generating pathways to Ras. GRF4 activates  
Ras in response to elevation of intracellular cAMP and/or cGMP. GRF4  
is a target for Nedd4 ubiquitination as it binds Nedd4. Activation of  
the Ras signaling pathway controls numerous cellular functions, such as  
cell metabolism, proliferation, differentiation and transformation.  
CC Therefore modulation of Ras activity may provide a mechanism for  
controlling diseases. GRF4 polynucleotides and polypeptides may be used  
in the treatment of diseases associated with inappropriate GRF4  
expression and activity such as cancers and neuronal disorders. The  
CC GRF4 polypeptides may be used as antigens in the production of  
antibodies against GRF4 and in assays to identify modulators (agonists  
and antagonists) of GRF4 expression and activity. The anti-GRF4  
CC antibodies and GRF4 antagonists may also be used to down regulate GRF4  
expression and activity. Inhibition of Ras can reduce cellulose  
proliferation and cancers.  
CC  
XX  
XX Sequence 1499 AA:  
Query Match 100.0%; Score 7721; DB 21; Length 1499;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1499; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKPLAIPANHGVMGOQEHSLPADFTKLHLTDSLHPQVTHVSSSHSGCSTISDSSSSLS 60  
DB 1 MKPLAIPANHGVMGOQEHSLPADFTKLHLTDSLHPQVTHVSSSHSGCSTISDSSSSLS 60  
QY 61 DIYATESEAGDMDLSGLPETAVDSEDDDEDIERASDPLMSRDIVDCLEKDPIDRTD 120  
DB 61 DIYATESEAGDMDLSGLPETAVDSEDDDEDIERASDPLMSRDIVDCLEKDPIDRTD 120  
QY 121 DDIFQLLEFMHQLPAFANMTWSVRELCAVVFVAVVERAGTIVLNDGEELDSWVILNGS 180  
DB 121 DDIFQLLEFMHQLPAFANMTWSVRELCAVVFVAVVERAGTIVLNDGEELDSWVILNGS 180  
QY 181 VEVVYVPGKABILLCMGNSFGVSPMDKRYMKGVMTKVDCCQFCIAQODCCRIILNQEK 240  
DB 181 VEVVYVPGKABILLCMGNSFGVSPMDKRYMKGVMTKVDCCQFCIAQODCCRIILNQEK 240  
QY 241 NMQVVEEGELVYMKENHRELDRTGRKHIVIKGTSERLTMHLEHNVVPTFIEDLL 300  
DB 241 NMQVVEEGELVYMKENHRELDRTGRKHIVIKGTSERLTMHLEHNVVPTFIEDLL 300  
QY 301 TYRFLSSPMVEGKLLLEWNPDLRDKVTRVVLVNNHNEDEGDPAAMTRFLEEFPNN 360  
DB 301 TYRFLSSPMVEGKLLLEWNPDLRDKVTRVVLVNNHNEDEGDPAAMTRFLEEFPNN 360  
QY 361 LERKMGHLLRLNIACAAKAKKRLMTLTKRSRARPFLFILLGSEKGFIVDSVDSGS 420  
DB 361 LERKMGHLLRLNIACAAKAKKRLMTLTKRSRARPFLFILLGSEKGFIVDSVDSGS 420  
QY 421 KATEAGLKGDOILEVNGONFENIQLSKAMEILLNNTHLSITVKNLDFVEKELLTRLSEE 480  
DB 421 KATEAGLKGDOILEVNGONFENIQLSKAMEILLNNTHLSITVKNLDFVEKELLTRLSEE 480  
QY 481 KRNGAPHLPKIGDIKAKSRISIDPLAVDEVOYIGLEVYKNSKANTYVGNRKLKIIDKT 540  
DB 481 KRNGAPHLPKIGDIKAKSRISIDPLAVDEVOYIGLEVYKNSKANTYVGNRKLKIIDKT 540  
QY 541 RISTLPQKPYNDIGIGSODSTVGLROTTHIIPALPVSGTLSSSNPDLLQSHHRIDFS 600  
DB 541 RISTLPQKPYNDIGIGSODSTVGLROTTHIIPALPVSGTLSSSNPDLLQSHHRIDFS 600  
QY 601 ATPDLPOVLKFKADQOSRYIMISKDTAKEVYIQAIREFAVATPDQYSLCEVSYTPE 660  
DB 601 ATPDLPOVLKFKADQOSRYIMISKDTAKEVYIQAIREFAVATPDQYSLCEVSYTPE 660  
QY 661 GVIKORRLPDQSLKADRIQLSGRYVYLLKNNMETETLCSDEDAQLRESQISLLQSTVE 720  
DB 661 GVIKORRLPDQSLKADRIQLSGRYVYLLKNNMETETLCSDEDAQLRESQISLLQSTVE 720  
QY 721 VATQLSKRNFELEFNIEPTXYIDLFKLRSKTSKANLKRFEVINOTEFVWASRIILETN 780  
DB 721 VATQLSKRNFELEFNIEPTXYIDLFKLRSKTSKANLKRFEVINOTEFVWASRIILETN 780  
QY 781 QLRKMKIIRKIKIALHCRECKNFNSMFAISGTLAPVARLRTWELPKNYEKLFDL 840  
DB 781 QLRKMKIIRKIKIALHCRECKNFNSMFAISGTLAPVARLRTWELPKNYEKLFDL 840  
QY 841 QDLFDPSSNNAKRYNVLSQNLQPIIPLFPVYIKKDLFLHEGDSYVDGLVNEKRLMI 900  
DB 841 QDLFDPSSNNAKRYNVLSQNLQPIIPLFPVYIKKDLFLHEGDSYVDGLVNEKRLMI 900  
QY 901 AKETRHVGRMASVMMDPALMFRTRKKKWSRSGSOSTNATVLDVOTGGHKRRVRS 960  
DB 901 AKETRHVGRMASVMMDPALMFRTRKKKWSRSGSOSTNATVLDVOTGGHKRRVRS 960  
QY 961 FLNAKKLYEDAQMARVKYQVLSNLELEMEDESLQTLSCCEPATNTLPKNFGDKRPVKE 1020  
DB 961 FLNAKKLYEDAQMARVKYQVLSNLELEMEDESLQTLSCCEPATNTLPKNFGDKRPVKE 1020  
QY 1021 TSPVAPRAGSQKQKQSLPQPPQOPPPAKHINOGLQVAVASLYPRKKVYPKVDLPFGINS 1080  
DB 1021 TSPVAPRAGSQKQKQSLPQPPQOPPPAKHINOGLQVAVASLYPRKKVYPKVDLPFGINS 1080  
QY 1081 PQALKIILSSEESLEPHKQKQADTISMASSQSSPPTSPQSSPRKGYTLAPSGTYDNF 1140  
DB 1081 PQALKIILSSEESLEPHKQKQADTISMASSQSSPPTSPQSSPRKGYTLAPSGTYDNF 1140  
QY 1141 SDSGHSSEISRSIVSNSSPDSVPSLHDERORHSYIVETNLGMGMERRTIMIEPDQY 1200  
DB 1141 SDSGHSSEISRSIVSNSSPDSVPSLHDERORHSYIVETNLGMGMERRTIMIEPDQY 1200  
QY 1201 SLGSAFMSSEGRGLYATATVYSSPSTEBLSODQCDRASLDAADSGRWSVSCSSGSHDNI 1260  
DB 1201 SLGSAFMSSEGRGLYATATVYSSPSTEBLSODQCDRASLDAADSGRWSVSCSSGSHDNI 1260  
QY 1261 OTIOHORSWETLPPGHHTFDYSGDPAGLMASSIMDOIMFSDHSTKYNROMSRESLEOA 1320  
DB 1261 OTIOHORSWETLPPGHHTFDYSGDPAGLMASSIMDOIMFSDHSTKYNROMSRESLEOA 1320  
QY 1321 QSRASMASSTGYWGEDSEGDGTIKRRGKQVSYLEAESSLTSTYETETKVPMPAHIAV 1380  
DB 1321 QSRASMASSTGYWGEDSEGDGTIKRRGKQVSYLEAESSLTSTYETETKVPMPAHIAV 1380  
QY 1381 ASSTTKGLIARKEGRYRPPPTPGYIGIPITDFPEGHSHPARKPPDYNAVALQSRVAV 1440  
DB 1381 ASSTTKGLIARKEGRYRPPPTPGYIGIPITDFPEGHSHPARKPPDYNAVALQSRVAV 1440  
QY 1441 SSDPAGSSVQOPGHGHTSSRPVKNKPMWKNPNESDPLAPVQSGFSEDEDEQVSAV 1499  
DB 1441 SSDPAGSSVQOPGHGHTSSRPVKNKPMWKNPNESDPLAPVQSGFSEDEDEQVSAV 1499

RESULT 2  
AAB42658  
ID AAB42658 standard; Protein; 1675 AA.  
XX  
XX AAB42658;  
AC  
XX  
XX 08-FEB-2001 (first entry)  
DT  
XX  
XX Human ORFX ORF2422 polypeptide sequence SEQ ID NO:4844.  
DE  
XX  
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulerary; antiprosclitic; antiparkinsonian; noctropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antineoplastic;  
KW antiviral; antibacterial; antifungal; antineumatic; antihypoid;  
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200058473-A2.  
PN  
XX  
XX 05-OCT-2000.  
PD  
XX  
XX 31-MAR-2000; 2000WO-US08621.  
PE  
XX  
XX 31-MAR-1999; 9905-0127607.  
PR  
XX 02-APR-1999; 9905-0127636.  
PR 05-APR-1999; 9905-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX  
XX (CURA-) CUBAGEN CORP.  
PA  
XX  
XX Shinkets RA, Leach M;  
PI  
XX  
XX MPI: 2000-602362/57.  
DR N-PSDB; AAC76867.  
DR  
XX  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
XX  
XX Claim 11: Page 4028-4032; 5507pp; English.  
XX  
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytosolic; hepatotropic; vulerary;  
CC antiprosclitic; antiparkinsonian; noctropic; neuroprotective;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;  
CC antihypoid; and antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX  
SQ Sequence 1675 AA;  
Query Match 60.2%; Score 4645; DB 21; Length 1675;  
Best Local Similarity 64.3%; Pred. No. 0;  
Matches 962; Conservative 163; Mismatches 296; Indels 76; Gaps 26;  
QY 20 SLPADFTKLHLPDLSLPOVTHVSSSGSITSDSSSSLSLDIYQATESPAGMDLSGLP 79  
DB 238 SLPADLTLMHLENPHQVTHVSSSGSITASDSSSSLSLDIYQATESPAGMDLTRLP 237  
QY 80 ETAVDESEDDDE-EDIERASDPLMSRDIVRDCLEKDPIDRTDDIDQLLEFHMQLPAFAN 138  
DB 298 EGVDESEDEDEDELEIDR-TDPIQGRDIAVECELEKPEADKTDIDQLLEFHMQLPAFAN 356  
QY 139 MTMSVRELCANVFAVERAGTIVLNDGEELDSWYILNGSEVYYPDGAELICGNS 198  
DB 357 MTMSVRELCANVIFEVDEAGAIILEDDGELDSWYIILNGVEISHPDGKVENLEPGNS 416  
QY 199 FGVSPTMDREYMGVMRTKVDCQFVCIAQODYCRILNOYKMKQVEEGELVMYKEHR 258  
DB 417 FGITPTLDKQYMHGIVRTKVDCCQFVCIAQODYWRILNHKMTHKVEEGELVMYHEHR 476  
QY 259 ELDRGTGRKHIVIKGTSERLIMHVEHSVDPFTIEDFLTYRFLSSPMVEYKKLLE 318  
DB 477 ELDRSGTRKHIVIKATPEKLMHLEHSIVDPTYIEDFLTYRFLLESPLDVGIKLE 536  
QY 319 WENDPSLRDKVTRVAVLLVNNHFNDEBGPAMTRFLEEFENNLEREKMGHLRLNATCA 378  
DB 537 WFKIDSIRDKVTRIYLLVNNHFNDEBGPAMTRFLEEFENNLEDTKMGNHLRLNATCA 596  
QY 379 AKAKRLMTLTRSREAPLRFILLGSEKGEFIVDSVDSKATEGKLRGQIILEVNG 438  
DB 597 AKAKRQVNLQKASRESPLQFSLNGSEKGEFIVDEBEGSKRADSGKLRGQIMVNG 656  
QY 439 QNFENIOLSKAMEILRNNTLSTVKTNLVFVKELTLTRLEEKRNCAPIHPIKIDIKAS 498  
DB 657 QNFENITMKAVELRNNTLALTVKNIFVKELFLR-TEGKSGVPHIPIKIAE-KKSN 714  
QY 499 RYSIPDLAVDEQVIGLEKVNKKSKANTV-GGRNKKLLIKDKTRISILPQKPYNDIGIQ 557  
DB 715 RHSIOHVPGDIEQT-SEKSKKKKANTVSGGRNKKIRIKIDKTRFSLPKFLSDGSLQ 773  
QY 558 SODDSIVGLRQTHIPALVSGTSSNPDLQSHHRILDFKATPLPROVLRVFAQD 617  
DB 774 SODDSIVGTGRCHSLAIMPICGLTSSSPDLQPTTSMDFSNPDIPOQVLRVFAVDQ 833  
QY 618 QSRVIMSKDPTAKKEVYIQAIRFAVATPDQYSLCEVYTPBEGVIRKQRLPQDLSKILAD 677  
DB 834 QSCYIIISKDTAKEVYFAVHVEGGLGASDTYSLCEVYTPBEGVIRKQRLPQDLSKILAD 893  
QY 678 RIOLSGRYVYKNNMETTLCSDSDAQLRESQISLQISTVEVATOLSKRNEELFNIE 737  
DB 894 RIOLNGRYVYKNNMETTLCSDSDAQLRESQISLQISTVEVATOLSKRNEELFNIE 953  
QY 738 PTEYIDLEFLRKSQTSCANLKREFEVINQETFWVASILLIETNOLKRMKIKIKHRIKIALH 797  
DB 954 PTEYIDLEFLKNSKTGTHTLKRFEDIVNOETFWVASILLIETNOLKRMKIKIKHRIKIALH 1013  
QY 798 CRECKNFNSFAFIISGNTLAPVARLRTTWELPKRYKELTQDLODLDPDPSRNMAKYNNVL 857  
DB 1014 CRECKNFNSFAFIISGNTLAPVARLRTTWELPKRYKELTQDLODLDPDPSRNMAKYNNVL 1073  
QY 858 NSONLQPIITLPEVPIKRDLTFLHEGNDKVDGLVNEFKLMIAKETIRHVGMAVNDP 917  
DB 1074 SSQSMQPIITLPEVPIKRDLTFLHEGNDKVDGLVNEFKLMIAKETIRHVGMAVNDP 1133  
QY 918 ALMFRTRKKKWRSLSGSOSTNATVLDVAQTGCHKRRVRSSTLNAKKLYEDAQMARKV 977  
DB 1134 AMMFROR-----SLQSGSTNSNMIDV-QGGAKKKRRARRSSLLNAKKLYEDAQMARKV 1184  
QY 978 KOYLSNLELMEDEESLQTLSDCEPATNTLPKNGCDKPKV--SETSPVAPRAGSQQKQ 1035  
|||||::: ||| |::||| ||| |||:: | ||||| |:: |

```
Db 1185 KQYLSLDVETDEEKFQMMSLQMEPAYGTLTKNLSEKRAKXSSEMSFVPMRSAGOTTTKA 1244
QY 1036 SLRPOQGPAPRAKINGLOVAPVSLYPRKKVYKDLPRPGINSPOALKKILSLSEGS 1095
Db 1245 HLHQP-----HKVSOVLQVAVNLHPIRKKGOTKD--PALNTSLPQ---KVLSTTEIS 1293
QY 1096 LERHKQAEPTISNASSOLSSPSPKGYTLAPSGTVDNFSPGSHSEISSRSIV 1155
Db 1294 ---GKHTEDTISVASSLHSSPPASPGSPHKGYTLIPSAKSDLSDSSHSSEISSRSIV 1350
QY 1156 SNSFSFVPSVLIHERQORHSVSLVENLGMGRMERTMTPEPDYSLGSIAPMSEG---- 1211
Db 1351 SNCSVDMSALQDERSSQALAVPESTGALTEKTEHAS-----GIGDHSQHGPGWTL 1403
QY 1212 -RGLYATATVISPSTEELSDQGDGRASLDPAADSGRSMWTSSCSSGSHDNIOTIOHRSWE 1270
Db 1404 KPSLIKCLAVSSSVSNEISQ---EHIIIEADGRCRSMWTSSCSSSHDNPOLNPKRSMW 1460
QY 1271 TL-PFGHTHEDYSGDPAGLWASSHMDQIMFSHSTYKNNQNSRESLEQAQSRASWASS 1329
Db 1461 FLNSYRHTHLD---DPI---AEVEPTDSEPYSCSKSCSRICGCKGSLF---RKSMTSS 1510
QY 1330 TGYWGESESGDTGIRKRGKADVISEASSLSTVTEETKRPVMPAHIVASTTGGLI 1399
Db 1511 SSL-SDTYEPNYGTVKR-----VLESTPAESSEGLDPKDATDPYKTYVSTSTERGLI 1562
QY 1390 A-----RKGRARYREPPPTPGYIGIPITDFPEG-HSHPARKPPDYVVALQSRHVARSS 1442
Db 1563 VYCVTSRKDDRYREPPPTPGYIGISLADLKEGPHTH--LKPPDYSAVAQRSMHNSL 1620
QY 1443 DTAGPSSVQCPHGHTSSRYNKRQMHKPNESDRLAPYOSQGSTEEDEDEOVSAV 1499
Db 1621 SRUPPASLSSNLACVSPKIVTQPQRHNLQPFHKPLGDVTD--DSEADENEQVSAV 1675

RESULT 3
ABG66725
ID ABG66725 standard; Protein; 1651 AA.
AC ABG66725;
XX
XX 30-AUG-2002 (first entry)
XX
XX Human novel polypeptide #60.
XX
XX Human: inflammatory condition; shock; sepsis; immune response;
XX cancer; wound healing; central nervous system disease; haematopoiesis;
XX peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
XX myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
XX cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
XX bone degenerative disorder; periodontal disease; reperfusion injury;
XX lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
XX allergic condition; thrombolytic; thrombosis; coagulation disorder;
XX fungal infection.
XX
XX Homo sapiens.
XX
XX WO200244340-A2.
XX
XX 06-JUN-2002.
XX
XX 30-NOV-2001: 2001WO-US47004.
XX
XX 30-NOV-2000: 2000US-0028952.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Goodrich RM, Liu C, Zhou P, Asundi V, Wang J, Wang D;
XX Yamazaki V, Ujwal ML, Drmanac RT;
XX
XX WPI; 2002-508509/54.
XX
XX N-PSDB; ABR94949.
XX
```

```
PT Novel nucleic acids and polypeptides for diagnosis, treatment of
PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT disorders, cancer and promoting wound healing.
XX
XX Claim 10: Page 629-632; 672pp; English.
XX
CC The invention relates to human novel polynucleotides and associated
CC polypeptides. The polynucleotides and polypeptides are useful for
CC treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC and cancer and for promoting wound healing. The sequences are used to
CC induce the proliferation of neural cells and regeneration of nerve and
CC brain tissue, and are useful for the treatment of central and peripheral
CC nervous system diseases and neuropathies, such as Alzheimer's disease,
CC Parkinson's disease, Huntington's disease and amyotrophic lateral
CC sclerosis. The sequences are involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
CC cell disorders and platelet disorders such as thrombocytopenia,
CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC disease. The sequences of the invention are also useful for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, immune deficiencies and disorders
CC including severe combined immunodeficiency (SCID), bacterial or fungal
CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC gravis, allergic conditions such as asthma, thrombolytic or thrombotic
CC novel polypeptides of the invention.
XX
XX Sequence 1651 AA:
XX
XX Query Match 60.2%; Score 4644.5; DB 23; Length 1651;
XX Best Local Similarity 64.3%; Pred. No. 0;
XX Matches 962; Conservative 163; Mismatches 296; Indels 75; Gaps 26;
XX
QY 20 SLRADFTKHLTDSLAPQVTHVSSSHSGSITSDSGSSLSDIYQATESPAGMDLSGLP 79
Db 215 SLRADLTKMLTENPHNPQVTHVSSSGCSIASDSGSSLSDIYQATESPAGVDLTLRLP 274
QY 80 ETAVDSRDDDD--EDIERASDPLMSRDIVYDCLKRPIDTDDIDQLLEFHMQLPAFAN 138
Db 275 EGPVDSDEDEDEEIDR--TDPLQGRDLVRECLEKEPADTDDIDQLLEFHMQLPAFAN 333
QY 139 MTMSVRELCVAVWFAVERAGTIVLNDGEELDSWVYILNGSEVAVYPODKAEILCMGNS 198
Db 334 MTMSVRELCVWIFEVYEDAGALILEDEGLDSWVYILNGYIEISHPGKVENLWNGNS 393
QY 199 FGVSPTMDKRYMKGMVTKVDDCOFYCIAQODYCRILINOVKMKQVKEEGETVMVKEHR 258
Db 394 FGITPTLDKQYMHGIVATKVDCCQFVCIAQODYWRIILNHVKMTNHHVEEGETVMVNEHR 453
QY 259 ELRBTGRKHIVYIKGSERLTMHIVEHSVVDPTFEDLTLVYRFLSPMGVKKILE 318
Db 454 ELDRSGTRKHIVYIKATPERLIMHLIEHSIVDPTTYIEDLTLVYRFLSPMDVGLKLE 513
QY 319 WFNDPSLRDKVTRVVLVWVNNHFNDEGDPAMTRFLEEFENNERKMGCHRLINIAQA 378
Db 514 WFKIDSLRDKVTRVTLVWVNNHFNDEGDPAMTRFLEEFENNERKMGCHRLINIAQA 573
QY 379 AKAKRRLMTITKPSREAPDLFIILGSEKGFIFVDSVDSGKATAGLKRGDQILEVNG 438
Db 574 AKAKRWQVVLQKASRSPLOFSLNGSEKGFIFEGVEGSKAASGKLKRGDQIMEVNG 633
QY 439 QNFENIOLSKAMEILRNNTLSTIVKTNLFVEKELLTRLUSEEKRNGAPHLPKIGDKKKS 498
Db 634 QNFENITPFKAVELIKNNHLLATVKTINLFVEKELLR--TEQKSSVPHPHPIAE--KKSN 691
QY 499 RYSIPDLAVDVOVIGLEKVNKSKANTV--GGRNKLKLLIKTRISILPOKPYNDIGIQ 557
Db 692 RHSIQHVPGDIEQT--SQEGSKVKVANTVSGGRNKRILKILDKRIFSLRPKLFSDGSLQ 750
QY 558 SQDDSIIVGLRQTKHIFPALPVSGTSLSSNPDLQSHHRIILDSAPDLPDQVLRVFKAQD 617
```

Db	751	SODSYICTGRCHRSLA	IMPIREPTLESSSPDLLOPTTSMIDPNSDLPDYOIRKFKDQ	810
Qy	618	QSRVIMISKDTAKEV	IOAIRFAVATAPDOYSLCEVSVPEGVIKORRLPDQLSLAD	677
Db	811	QSCYIISKDTAKEV	FAVHVEFGLTGASDPYSLCEVSVPEGVIKORRLPDQSKLAD	870
Qy	678	RIOISGRYIKNNMET	ETTLCSDEDAQELLRESIISLLOLSTVEVATQISMREFELRNIE	737
Db	871	RIOINGRYIKNNMET	ETTLCSDEDAQELVESOLSMLOLSTIEVATOLSMRDFELRNIE	930
Qy	738	PTEYIDDLFKLRST	CSANKFREEVYNQETFWASEIRRETNOLRKRIIKHFKIALH	797
Db	931	PTEYIDDLFKLNST	GTWTHLKRREDIYNQETFWASEILTEANOLKRIIKHFKIALH	990
Qy	798	CRECKNFNSMFAIS	IGLNLAVARLRTTEKLPNKYEKLFODLLODLFDSRMMAKYRNL	857
Db	991	CRECKNFNSMFAIS	IGLNLAVARLRTTEKLPNKYEKHLQDLODLOIPDSRMMAKYRNL	1050
Qy	858	NSQNLQPPILP	FLVYIKKDLTFLEHGNDKVDGLVNEFKLRMAIKELRHVGMASVNDP	917
Db	1051	SSOSMOPRIPLP	FLPVYKKDMTFLHEGNDKVDGLVNEFKLRMAIKELRHVGMATSANNDP	1110
Qy	918	ALMFRTRKKKRRSL	SGSOSTATAYLDVAQTSQGHKKRRVRSBSFLNAKKLYEDDAQMARKV	977
Db	1111	AMFRPRR-----	SLQSOSTSNMIDLV-OGGAHKKRRARRSSLNNAKKLYEDDAQMARKV	1161
Qy	978	KQYLSNLELEMEDE	SLTLOCEPATNTLPKPNQGGKRPVK-SETSPVAPRAGSOQKQAS	1036
Db	1162	KQYLSNLDVETDEK	FQMSLOMEPRAYGTLTKLSKRSKAKSEMSPVPMRAGOTTAH	1221
Qy	1037	LPOQOQPPRAH	KINOGLQVPAYSLYPSRRKRYPVKDLPPFGINSPOALKILSLSEEGSL	1096
Db	1222	LHQP-----	HRVSOVLQVPAYVNLHPIRKKQGTQD-PALNTSLDP--KVLGTETEIS-	1269
Qy	1097	ERHKQAEEDTSN	ASQSLSPPTSPQSPRKKGYTLAPSGVTNFPDSQSHSETSSKSTIYS	1156
Db	1270	--GKHTEDITSV	ASLSHSSPPASPOGSPHKGYTLPLPSAKSDNLSDSSHSEISSRSTIYS	1327
Qy	1157	NSFSDVSVSLHDE	RRORHVSIIYETNLGCMREERTMLERPOYSLGSYAPMSEG----	1211
Db	1328	NCSDVSMASMA	LODRCSQMLAVPESTGALEKTEHNAS-----GIGHDSQHGPGWTLIK	1380
Qy	1212	RGLYATATVIVIS	SPSTEELSDQDGRASLDAADSGRGSWTS CSGSHDNITQTOHQRSMET	1271
Db	1381	PSLJKCLAVSSV	SNEEISQ---EHIIIEAADGGRSMTSCSSSHDNFQSLPNPKSWDF	1437
Qy	1272	L-PGCHTHFYSG	PRADLMASSIMQIMFSHSTIKYNNQONSRESLEDAQSRASWAST	1330
Db	1438	LNSYRHHILD---	DPi---AEVEPTSEPSYSCSKSCSRCTGQCKSLD---RKSWTSSS	1487
Qy	1331	GYWGDESGDGT	IKRRGKGDVSIEMASSSLTAVTTEEMKRPVMPRAHIALAVASTTKGLIA	1390
Db	1488	SL-SDTEPRNT	GIYKKR-----VLESTPAESSEGLDRKDATDPYKTVIYSTTEKGLIV	1539
Qy	1391	-----RKGRYR	PRPTPGYIGITPTPERG-HSPARKPRDYVAVLQDRSMAVARSSD	1443
Db	1540	YCVISPKKDDYR	RRPPTPGYIGISLADLKEGRPHN--LKREDYSVAQRKMMHNLS	1597
Qy	1444	TAGPSSVOOP	GHPTSSRPVNAKPMKPNESDRLARLYOSQSGSTEEDQDEQVSAV	1499
Db	1598	RLPRASLSNL	VAVCPSPKIVTQQRHNLQFPHKLGADYDA--DSEADENEGVSAV	1651
RESULT 4				
ABB64459				
ID	ABB64459 standard; Protein; 1551 aa.			
xx				
AC	ABB64459;			
DT	26-MAR-2002 (first entry)			
xx				
DE	Drosophila melanogaster polypeptide seq ID NO 20169.			

[illegible]

```
QY 396 PLPEFILGSGE-----KGFGLFVDSVDSKATKTEAGLRGQDILEVNGCNFENIQ 445
    ||:|||||
    ||:|||||
Db 416 PLNFRIVGVELGVALATGNAVAIGIYISHVEFGSKAQDGLKRGQDIIHEVNGSLDHT 475
QY 446 LSKAMELNRNTHLSTVTKNLFVEFKELLTRLSEE-----KRGAPHLR 489
    ||:|||||
    ||:|||||
Db 476 SKRALELTGTHTLSISVKSNLLGFKEIMQALHGGGTAGSGSISAGSGSKSVRPRRI 535
QY 490 KIDDIRK-----AS-----RYSIPDLAV----- 508
    ||:|||||
    ||:|||||
Db 536 CANDIARLHGRSSTTDELSSVSASNRHAYRLSSVMDLLDQPCAPQTPPVSGSGNMA 555
QY 509 ---VEQYIGLEKYN---KRSKAN-----TVGQ-----RNKLKILDKTRISILPQK 548
    ||:|||||
    ||:|||||
Db 596 SNFMQOL--LQSVNNSAKKSGNSNDQODTGGFMTLAPKRRLQALAKML----- 647
QY 549 PYNDIGIGQSDSIVGL-----RQTKHIFLALPVSG---TLSSSNPDL 589
    ||:|||||
    ||:|||||
Db 648 -LKNQNHSSSLNDSPTLLNDPKSKLSAVSSCSSSTQSSINGCTVSGGRGLYQSQSNPDL 706
QY 590 -----LQSH-HR-----ILDESAT-----PDPDYLYRFEKA 615
    ||:|||||
    ||:|||||
Db 707 TSLNVDGSDAGNGGGRGLQVNTLMAHHRPSAATLTSTQSHLLPDYDHYLKYKA 766
QY 616 DQOSRYIMISKDTAKEVYIOAIRFEAVTATPPQYSLCEVSVPEGYIKORRLPDOLSKL 675
    ||:|||||
    ||:|||||
Db 767 DQCKYVLIKETTAEHVWMLTLQEFQIHDPSSNFSICEVSVGDGKVKRRRLPDQLONL 826
QY 676 ADRIQUSGRYYLKNMNETFTLQSDEDAQELLRSQSLQLQSTVEVATQUSMRNFEFLRN 755
    ||:|||||
    ||:|||||
Db 827 AERISFAARYLLKLNDSLEPLVDELALLEYRESNHFHLLNAYELIQLTLDDFAFRQ 886
QY 736 IETETVIDDFKLRSTSCANLKRFEVINOETFWYASELIRETNOLKRMKIIKHFKIA 795
    ||:|||||
    ||:|||||
Db 887 IESTEVYDELFEIERSKRGVPMLSKFAELVNRKFWVSEICAHNIVRRKIKYQKFIA 946
QY 796 LHCRCENFNSMFAIISGLNLAPVARTWELTPNKYEKLFODLDFEDPSRNMAYRN 855
    ||:|||||
    ||:|||||
Db 947 RHCKECRNFNSMFAIYSGIGHAVSRRLQRTWEKLPKSYQGLFDLODMPSRNMASYRQ 1006
QY 856 VLNSQNL-QPITPLFPVYIKKDLTFLEHGDSVYDGLVNEFKRMIAKEIRHGVASVN 914
    ||:|||||
    ||:|||||
Db 1007 LVSARELLAOPHPIIPYPIYKDLTFIHLGNDTVGDLINEKRLMAKEVRLTLHHCCSP 1066
QY 915 MD-----PALMFTRRKKKMSISLSOGSTNA---TVLDVQOTGCH---KKVARSSTFLNA 964
    ||:|||||
    ||:|||||
Db 1067 YDLSTLLELKGSFSPNALFSLNOMASQSMAAAGTV--IAANAGQATIKRRKSTAAPNP 1124
QY 965 KKLVEDAOMARKVQYLSNLELEMEDESLQTLSLQCEPA-----TNTLPK--- 1009
    ||:|||||
    ||:|||||
Db 1125 KMFEEAOMVRRKAYKALNSLAKILSDEDLHKFLECEPAHSGTYSYSISHGNTSHHSQGG 1184
QY 1010 -----NFGDKKPVKSETS-----PVARAGSQOK 1033
    ||:|||||
    ||:|||||
Db 1185 GSISGAGGSSGGGSSSLNAGDOLSIYHTSSAPNSSLKRRHPPSSPTLSTSS 1244
QY 1034 AOSLPPOOOPRAHKINGLOVPAVSIVSRKKVPVKDLPKPCINSPOALKKILSLEE 1093
    ||:|||||
    ||:|||||
Db 1245 TSSTSHQRR---QMHNNG-----PKFGTASPOAKVMKMLLSSES 1280
QY 1094 GSLERHKKAQEDTISNASSQLSS--PRTSPQSPRKGYTLAPSGTVDNFSDSGHSEISSR 1151
    ||:|||||
    ||:|||||
Db 1281 SKIRPH---QPEVPRHSGTMAGVIRPLHMHMAHGFSTPSPGVY-----TSRA 1326
QY 1152 SSTSVNSSFSDSVVSLHDERQRHNSIVETNLGMRMERKRMTEPD-----QYSLGSTA 1206
    ||:|||||
    ||:|||||
Db 1327 TSAVANVQCTSPSP-SPCSHRRLASGNNIIPRAIHERSHSDPAPAPPPLPVSYDLSLES-- 1383
QY 1207 PMSEGGGLATATAVISSPSTSELSODQGRASLDAAADSGSGSWTSSGSHNDIOTIQH 1266
    ||:|||||
    ||:|||||
Db 1384 -----SSVIT--FRDLPKRSVTS-----GSISGDSG-----YVHQ 1413
```

```
QY 1267 RSMETLFFGHTHFDDYSGDPAGLWASSHMDQIMFSDHSTKYNRQONSRESLEQAOGRASW 1326
    ||:|||||
    ||:|||||
Db 1414 QO-----HYHLOY-----QOQOQONSQHEPPEPY 1439
QY 1327 ASSTGWGEDSEDDTGTIKRRGKDVSIKAESSSLTSVTTEETKPYVMP---AHIAVAS 1382
    ||:|||||
    ||:|||||
Db 1440 TAA-----DCRLLOQISNNAVTRNL-NSPCOSTNTPSTPPPPPPQPTATLOLSA 1488
QY 1383 STTKGLARKKEGRYRE-----PPPPPPYIGIPIITDFPEG---HSHPAKPPDYANV 1430
    ||:|||||
    ||:|||||
Db 1489 PPTAAAYMHARSQOQLQOQOQSLAMPPEPPPY-NVP---PLGSIYSH----- 1533
QY 1431 ALQSRMVARSDDAGPSSVQCPGH 1456
    ||:|||||
    ||:|||||
Db 1534 -----HOGTASRHLNMHGH 1549

RESULT 5
AAW88983
ID AAW88983 standard; protein; 286 AA.
XX
AC AAW88983;
XX
DT 01-MAR-1999 (first entry)
XX
DE Polypeptide fragment encoded by gene 140.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; lissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal;
KW immune system; asthma; lymphocytic disease; brain; hepatitis; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; stenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
XX
PN MO9854963-A2.
XX
PD 10-DEC-1998.
XX
PF 04-JUN-1998; 98WO-US11422.
XX
PR 18-DEC-1997; 97US-0070923.
PR 06-JUN-1997; 97US-0048877.
PR 06-JUN-1997; 97US-0048881.
PR 06-JUN-1997; 97US-0048884.
PR 06-JUN-1997; 97US-0048893.
PR 06-JUN-1997; 97US-0048896.
PR 06-JUN-1997; 97US-0048899.
PR 06-JUN-1997; 97US-0048915.
PR 06-JUN-1997; 97US-0048949.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048972.
PR 06-JUN-1997; 97US-0049020.
PR 06-JUN-1997; 97US-0049375.
PR 05-SEP-1997; 97US-0057628.
PR 05-SEP-1997; 97US-0057635.
PR 05-SEP-1997; 97US-0057644.
PR 05-SEP-1997; 97US-0057647.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057661.
PR 05-SEP-1997; 97US-0057667.
PR 05-SEP-1997; 97US-0057761.
PR 05-SEP-1997; 97US-0057764.
PR 05-SEP-1997; 97US-0057770.
PR 05-SEP-1997; 97US-0057775.
PR 05-SEP-1997; 97US-0057778.
PR 06-JUN-1997; 97US-0048875.
PR 06-JUN-1997; 97US-0048878.
PR 06-JUN-1997; 97US-0048882.
PR 06-JUN-1997; 97US-0048885.
```

PR 06-JUN-1997: 9705-0048894.  
 PR 06-JUN-1997: 9705-0048897.  
 PR 06-JUN-1997: 9705-0048900.  
 PR 06-JUN-1997: 9705-0048916.  
 PR 06-JUN-1997: 9705-0048962.  
 PR 06-JUN-1997: 9705-0048970.  
 PR 06-JUN-1997: 9705-0048974.  
 PR 06-JUN-1997: 9705-0048973.  
 PR 05-SEP-1997: 9705-0057584.  
 PR 05-SEP-1997: 9705-0057629.  
 PR 05-SEP-1997: 9705-0057642.  
 PR 05-SEP-1997: 9705-0057645.  
 PR 05-SEP-1997: 9705-0057648.  
 PR 05-SEP-1997: 9705-0057651.  
 PR 05-SEP-1997: 9705-0057662.  
 PR 05-SEP-1997: 9705-0057668.  
 PR 05-SEP-1997: 9705-0057762.  
 PR 05-SEP-1997: 9705-0057765.  
 PR 05-SEP-1997: 9705-0057771.  
 PR 05-SEP-1997: 9705-0057776.  
 PR 06-JUN-1997: 9705-0048876.  
 PR 06-JUN-1997: 9705-0048880.  
 PR 06-JUN-1997: 9705-0048883.  
 PR 06-JUN-1997: 9705-0048892.  
 PR 06-JUN-1997: 9705-0048895.  
 PR 06-JUN-1997: 9705-0048898.  
 PR 06-JUN-1997: 9705-0048901.  
 PR 06-JUN-1997: 9705-0048917.  
 PR 06-JUN-1997: 9705-0048963.  
 PR 06-JUN-1997: 9705-0048971.  
 PR 06-JUN-1997: 9705-0049019.  
 PR 06-JUN-1997: 9705-0049374.  
 PR 05-SEP-1997: 9705-0057627.  
 PR 05-SEP-1997: 9705-0057634.  
 PR 05-SEP-1997: 9705-0057643.  
 PR 05-SEP-1997: 9705-0057646.  
 PR 05-SEP-1997: 9705-0057649.  
 PR 05-SEP-1997: 9705-0057654.  
 PR 05-SEP-1997: 9705-0057666.  
 PR 05-SEP-1997: 9705-0057760.  
 PR 05-SEP-1997: 9705-0057763.  
 PR 05-SEP-1997: 9705-0057769.  
 PR 05-SEP-1997: 9705-0057774.  
 PR 05-SEP-1997: 9705-0057777.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX  
 PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA;  
 PI Fan P, Feng P, Ferrie AM, Fischer CL, Florence C;  
 PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DM;  
 PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM;  
 PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;  
 DR WPI: 1999-059865/05.  
 DR N-PDB: AAV84550.  
 XX  
 PT New isolated human genes and the secreted polypeptides they encode -  
 PT useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders  
 XX  
 PS Disclosure: Page 108: 772pp: English.  
 XX  
 XX The invention relates to nucleic acid sequences (AAV84411 to AAV84633)  
 CC encoding human secreted proteins (AAW8534 to AAW88756). The secreted  
 CC protein gene sequences are deposited with the ATCC under deposit numbers  
 CC ATCC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,  
 CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host  
 CC cells comprising recombinant vectors containing the nucleic acid  
 CC sequences are used for the recombinant production of the secreted  
 CC proteins. The polynucleotide and amino acid sequences are useful for are  
 CC useful for preventing, treating or ameliorating medical conditions e.g.  
 CC by protein or gene therapy. Pathological conditions can be also  
 CC diagnosed by determining the amount of the new polypeptides in a sample

CC or by determining the presence of mutations in the new polynucleotides.  
 CC Specific uses are described for each of the polynucleotides, based on  
 CC which tissues they are most highly expressed in, and include developing  
 CC products for the diagnosis or treatment of cancer, neurodegenerative  
 CC disorders, developmental abnormalities and foetal deficiencies, blood  
 CC disorders, tumours, leukemias, diseases of the immune system, autoimmune  
 CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,  
 CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,  
 CC restenosis, prostate diseases, obesity, disorders involving osteoclasts  
 CC such as osteoporosis, arthritis or malignancies, diseases of testes,  
 CC lung or thymus, digestive/endocrine disorders, infections and AIDS. The  
 CC polypeptides are also useful for identifying their binding partners.  
 CC The present sequence represents a polypeptide fragment encoded by a  
 CC gene of the invention (see descriptor line for gene number).  
 XX  
 XX Sequence 286 AA:  
 SQ  
 Query Match 19.5%; Score 1508; DB 20; Length 286;  
 Best Local Similarity 99.3%; Pred. No. 2,7e-98;  
 Matches 284; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1214 LVATATVYISSPSTEEISQDQGDRAISLDADSGSGSWTSCSGSHDNIQTQHQRSMETLP 1273  
 DB 1 LVATATVYISSPSTEXLSQDQGDRAISLDADSGSGSWTSCSGSHDNIQTQHQRSMETLP 60  
 QY 1274 FGHTHFPGSDPAGLMASSSHMDQIMPSDHSKYRNQNSRESLEQNSRASMASSTGYW 1333  
 DB 61 FGHTHFPGSDPAGLMASSSHMDQIMPSDHSKYRNQNSRESLEQNSRASMASSTGYW 120  
 QY 1334 GEDSECDPTGTRKRGCKDVSIEASSSLTSTTEETRPVMPAHIAVASTTGILARKE 1393  
 DB 121 GEDSECDPTGTRKRGCKDVSIEASSSLTSTTEETRPVMPAHIAVASTTGILARKE 160  
 QY 1394 GRTRPPPTPPGYIGIPITDPFEGHSHARKPDYVNALQSRMVANSSPTAGPSSVQOP 1453  
 DB 181 GRTRPPPTPPGYIGIPITDPFEGHSHARKPDYVNALQSRMVANSSPTAGPSSVQOP 240  
 QY 1454 HGHTSSRPVKNKQWHPNCSDPRLAPYQSGFSTEDEDEQVSAV 1499  
 DB 241 HGHTSSRPVKNKQWHPNCSDPRLAPYQSGFSTEDEDEQVSAV 286  
 RESULT 6  
 ID ABB51059 standard; Protein: 286 AA.  
 XX  
 XX ABB51059;  
 DT 07-FEB-2002 (first entry)  
 XX  
 XX Human secreted protein encoded by gene 140 SEQ ID NO:1012.  
 XX  
 XX Human: secreted protein; immunomodulatory; antisclerotic; anti-HIV;  
 KW dermatological; immunosuppressive; antiinflammatory; immunostimulant;  
 KW cytostatic; cardiant; vascular; anti-angiogenic; ophthalmological;  
 KW neuroprotective; nootropic; anticonvulsant; antialzheimers; vinetary;  
 KW antiparkinsonian; antimicrobial; gene therapy; vaccine; immune disorder;  
 KW multiple sclerosis; systemic lupus erythematosus; HIV infection; cancer;  
 KW human immunodeficiency virus; hyperproliferative disorder; wound healing;  
 KW Gaucher's disease; cardiovascular disease; Schmitzer syndrome; chemotaxis;  
 KW Chaga's cardiomyopathy; coronary arteriosclerosis; angioecic disorder;  
 KW corneal graft neovascularisation; diabetic retinopathy; regeneration;  
 KW neurological disorder; Huntington's chorea; Alzheimer's disease;  
 KW Parkinson's disease; infectious disease; chromosome 4.  
 XX  
 XX Homo sapiens.  
 PN W0200162891-A2.  
 XX  
 XX 30-AUG-2001.  
 PD  
 XX  
 XX 21-FEB-2001; 2001WO-US05614.  
 XX

PR		24-FEB-2000:	2000US-184836P.
PR		29-MAR-2000:	2000US-193170P.
XX		(HUMA-) HUMAN GENOME SCI INC.	
PA			
XX			
PI	Ni J,	Ehner R, Lafleur DM, Moore PA, Olsen HS, Rosen CA;	
PI	Ruben SM, Soppet DR, Young PE, Shi Y, Florence RA, Wei Y;		
PI	Florence C, Hu J, Li Y, Kyaw H, Fischer CL, Farlie AM, Fan P;		
PI	Feng P, Endress GA, Dillon PJ, Carter KC, Brewer LA, Yu G;		
PI	Zeng Z, Greene JM;		
DR	WPI; 2001-625724/72.		
XX			
PT	Nucleic acids encoding 207 human secreted polypeptides, useful for preventing, diagnosing and/or treating, e.g. cancers, Parkinson's disease and diabetic retinopathy -		
PS	Disclosure: Page 313; 1533pp; English.		
XX			
CC	AB50301 to AB51287 and ABA83194 to ABA83441 represent human secreted proteins (I) and polynucleotide (II) sequences. (I) and (II) have various activities based on the tissues and cells the genes are expressed in.		
CC	Example of these activities include: immunomodulatory; anticlerotic; dermatological; immunosuppressive; antiinflammatory; immunostimulant; anti-HIV; cytotoxic; cardiant; anti-angiogenic; ophthalmological; neuroprotective; nootropic; anticonvulsant; antialzheimers; vascular; antiparasitolan; antimicrobial; and vulnerrary. (I) and (II) can be used in gene therapy and vaccine production. (I) and (II) can be used in the prevention, diagnosis and treatment of immune disorders (e.g. multiple sclerosis), systemic lupus erythematosus and human immunodeficiency virus (HIV) infections), hyperproliferative disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases (e.g. scintilar syndrome, Chagas' cardiomyopathy and coronary arteriosclerosis), angiogenic disorders (e.g. corneal graft neovascularisation and diabetic retinopathy), neurological disorders (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease), infectious diseases and/or for promoting wound healing, regeneration and/or chemotaxis. ABA83185 to ABA83193 and AB50300 represent sequences used in the exemplification of the present invention.		
CC			
XX			
SQ	Sequence	286 AA;	
	Query Match	19.5%;	Score 1508; DB 22; Length 286;
	Best Local Similarity	99.3%;	Pred. No. 2.7e-98;
	Matches 284;	Conservative 0;	Mismatches 2; Indels 0; Gaps 0;
QY	1214 LVATAVISSPSTEELISQDGDGRASLDADSGGWSVSCSSGGHDNIQTIOHRSMETLP	1273	
DB	1 LVATAVIASSPSTEXLSIQDDGRASLDADSGGWSVSCSSGGHDNIQTIOHRSMETLP	60	
QY	1274 FGHTHPFYSDDPGGLMASSSHMDOIMPSDSHTXYNRONOSRESLEAOASFASWASTGYW	1333	
DB	61 FGHTHFYSDDPGLMASSSHMDOIMFSDSHTXINRONORESLEAOASFASWASTGYW	120	
QY	1334 GEDESGDTGIKRRGKDVSIEAESSLSVTTEETKPVMPAHIAVASSTTGLIARKE	1393	
DB	121 GEDESGDTGIKRRGKDVSIEAESSLSVTTEETKPVMPAHIAVASSTTGLIARKE	180	
QY	1394 GRYPEPPPTPEGTIGITTPDPBECHSPARKPPDYNALDRSMVAASSDTAGPSSVOOP	1453	
DB	181 GRYPEPPPTPEGTIGITTPDPBECHSPARKPPDYNALDRSRKRVANSSTAGPSSVOOP	240	
QY	1454 HGHTSSRPVNKPOMHKPNESDPLAYQSOGFTEDEDEDOVSAY	1499	
DB	241 HGHTSSRPVNKPOMHKKNESDPLALAYQSOGFTEDEDEDOVSAY	286	
RESULT 7			
AAV70964			
ID	AAV70964 standard; Protein: 884 AA.		
AC	AAV70964;		
XX			

09-AUG-2000 (first entry)

Rat Ras signalling pathway associated protein *cAMP-GEFI*.

Rat: Ras signalling pathway; *cAMP-GEFI*; cyclic adenosine monophosphate; GEF; guanine nucleotide exchange factor; Rap1a; diagnosis; treatment; *cAMP-GEF*-associated disorder; drug; transgenic animal model; Ras-associated cancer; protein therapy.

*Rattus norvegicus*.

Key

Binding-site

Location/Qualifiers

231..300

/label= *cAMP-binding\_domain*

278..282

/note= "conserved motif in the *cAMP* binding pocket; amino acid at position 280 determines binding specificity for *cAMP/cGMP*"

619..642

/label= *SCR1*

/note= "Structurally conserved region which is highly homologous to Ras-superfamily GEFs"

692..734

/label= *SCR2*

/note= "Structurally conserved region which is highly homologous to Ras-superfamily GEFs"

771..792

/label= *SCR3*

/note= "Structurally conserved region which is highly homologous to Ras-superfamily GEFs"

WO200024766-A2.

04-MAY-2000.

22-OCT-1999; 99WO-US24826.

23-OCT-1998; 98US-0105507.

16-NOV-1998; 98US-0108685.

(MAST ) MASSACHUSETTS INST TECHNOLOGY.

Kawasaki H, Graybiel A, Housman D;

WPI: 2000-350690/30.

N-PSDB: AAD00314.

Isolated nucleic acid comprises nucleotide sequence encoding protein selected from normal or mutant *Ca*ldag-guanine nucleotide exchange factor -

Claim 64; Page 107-109; 128pp; English.

The present sequence is a cyclic adenosine 3', 5' monophosphate-guanine nucleotide exchange factor1 (*cAMP-GEFI*) from rat. The *cAMP-GEFI* has *cAMP* binding domain and Ras superfamily GEF domains. It has substrate specificity for Rap1a and is differentially distributed in brain and various other tissues. It selectively activates the Ras superfamily small G protein and Rap1a, functions as a negative regulator of Ras and directly couples the *cAMP* signal transduction system to Ras superfamily cascades. The present sequence is used for diagnosis and treatment (protein therapy) of *cAMP-GEF*-associated disorders. It is also useful for developing drugs and producing cell-lines or transgenic animal models for Ras-associated cancers.

Sequence 884 AA;

Query Match 8.8%; Score 676; DB 21; Length 884;

Best Local Similarity 23.9%; Pfad. No. 1.2e-38;

Matches 240; Conservative 143; Mismatches 277; Indels 344; Gaps 32.

79 PETAVSEDDDDDEDDIE-----RASDPLMSNDIYADCLEKPIRDTDDDIQLELFWM 131



```

Db      145 PEPQAGTHDVEEELVEAMALLSQGRDALLT-----VALRKSPGQRTDEILDLEELV 199
QY      132 QLFAPAMTMSVARELCAYVAVFVVERAGTIVINDGEELDSVILNGSVE-VTYPDGA 190
Db      200 HIAVAHLSSVSVARELAAYVLLFEPHSHKAGVLFPSQDGKGMWYIKNGSVNVTTRKGGV 259
QY      191 EILCMGNSFVSPTMDKEVYKGMRTKVDCCFVCIADOCYCIILNQVEKNMGKEEGE 250
Db      260 TTHEDDDGQALVADAPRAATIIIRENNCHFLRYDKODFNRIIDVEAKYMLREHGK 319
QY      251 IYVWKEHRELDRTGT-----RKGHIVIKGTSERLTMHVE---EHSVDP--T 293
Db      320 VVLY-----LERTSQAGSPRPPTPGRNRYTVMSGTEPKILELLFAMRDSANDPTET 374
QY      294 FIDFLLTYRTFLISSPMVEGKILLENFNDPS-----LRDKYTRVLLWY 337
Db      375 FLSDFLTLHSVFMPCTQLFALLHNFHVEPSEBAGSEDERSTYICNKRQOILRLVSRW 434
QY      338 NMHFNDEGDPAMTRFLEEPE-----NMLERE---KMGGHLRLNTAC---AAKAK 382
Db      435 ALXSPMLRSPPVATSFLOKISDLVSRDTRLSNLRREOYPPRRHHLEN-GCGNVSPQTK 493
QY      383 RLMLTLTKPSREAPLRFILLGSEKGFGLFVDSVDSGSKATEAGLRGQDILEVNGONEE 442
Db      494 ARNAPVWFRNHEERLP-----SSAGAIRVGDKV-----521
QY      443 NIDLSAMELIRNN-----THLSITYKTNLFVKELLTRISEKRNKGAIRLPKIGDIKKA 497
Db      522 ----PYDICRBDHSLVTLHLPTASV-----REYMAALANE---DHMTK-----558
QY      498 SRSYIDPLADVIOVIGLEKVNKSKANFVGGNKLKILDKRISILPKPYNDIGIGQ 557
Db      559 ----GOVL-----VKVNSAG-----569
QY      558 SODDSIVGLKQTKHPIPTALPVSGTLSSNPDLQSHRILDFSGATPDLPDQVLRVKADQ 617
Db      570 ----DVYGLQ-----PDARGVATSLGNE-----RIFVDP 596
QY      618 QSRVIMSKDTAKEVYIAQIRFPAVATPDPQYSLCEVSVTPGCVIKORRLPQLSKAD 677
Db      597 QEVH-----ELTPH-----PEQLG-----610
QY      678 RIQLSGRYYLKNMNETETLCSDAQELLRESQISLLQ--STEVAVATQLSMRNFELEFNI 736
Db      611 ----PTLGSSE-----MLDLVSAKDLGQLTEHDMNLEFNI 642
QY      737 EPTETIDLEF--KLRSKTSKANLKREEVINOETFWVASEILRETNOLKRMKIIKHFIK 793
Db      643 HOVELLHYVLGPQHLRDVTT-ANLEFRMRFNELQYVWATLELCVPPGRADLLKRFIK 701
QY      794 IALHCECKKFNFMFALISGLNLAPVARLKTWEKLPNKYEKLPDQLQDLFEDSRNNAKY 853
Db      702 LAHLKEOKKMLNSFFAVMEGLSNSAISRLAHATERLPHKRYKLYSALERLLDSMNRVY 761
QY      854 RNVLNSONLOPPIIPLFVPIKDLTFLHEGNSKVGLVFEKLMIAKIRIVGRASV 913
Db      762 RLALT--KLSPVIFPMPDLKDMTFLHESNHTLVEHLFEEMRMARAVRLHHCRSH 819
QY      914 NMDPALMFTTRKKKWSLGSLSOGSTNAVLDVAOTGHHKRRRSSFLNAKLLEADQAM 973
Db      820 STAPLSPLSR-----VSHIHDSQA 840
QY      974 AR-----KVKQYISNLELDEDSLOTLQCEP 1002
Db      841 SRISTCSEOSLSTRSPASTWAVYQOLKVIDNORELSRLSRELEP 884

```

```

XX      09-AUG-2000 (first entry)
DT
XX      Human Ras signalling pathway associated protein CAMP-GEFI.
DE
XX      Human: Ras signalling pathway: CAMP-GEFI; cyclic adenosine monophosphate;
KW      GEF; guanine nucleotide exchange factor; Rap1A; diagnosis; treatment;
KW      CAMP-GEF-associated disorder; drug; transgenic animal model;
KW      Ras-associated cancer; protein therapy.
XX
OS      Homo sapiens.
XX
XX      Key
FH      Binding-site
FT      231..300
FT      /label= CAMP-binding_domain
FT      278..282
FT      /note="conserved motif in the CAMP binding pocket;
FT      amino acid at position 280 determines binding
FT      specificity for CAMP/cgmp"
FT      616..639
FT      /label= SCR1
FT      /note="Structurally conserved region which is
FT      highly homologous to Ras-superfamily GEFs"
FT      689..731
FT      /label= SCR2
FT      /note="Structurally conserved region which is
FT      highly homologous to Ras-superfamily GEFs"
FT      768..789
FT      /label= SCR3
FT      /note="Structurally conserved region which is
FT      highly homologous to Ras-superfamily GEFs"
XX
PN      WO200024768-A2.
XX
PD      04-MAY-2000.
XX
XX      22-OCT-1999; 99WO-US24826.
XX
PF      23-OCT-1998; 98US-0105507.
PR      16-NOV-1998; 98US-0108685.
XX
PA      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
PI      Kawasaki H, Graybiel A, Housman D;
XX
DR      MPI; 2000-350690/30.
XX
DR      N-PSDB; AAD00315.
XX
PT      Isolated nucleic acid comprises nucleotide sequence encoding protein
PT      selected from normal or mutant CalDAG-guanine nucleotide exchange
PT      factor -
XX
XX      Claim 64; Page 113-115; 128pp; English.
XX
XX      The present sequence is a human cyclic adenosine 3', 5',
XX      monophosphate-guanine nucleotide exchange factor1 (CAMP-GEFI) which
XX      has CAMP binding domain and Ras superfamily GEF
XX      domains. It has substrate specificity for Rap1A and is differentially
XX      distributed in brain and various other tissues. It selectively activates
XX      the Ras superfamily small G protein and Rap1A, functions as a negative
XX      regulator of Ras and directly couples the CAMP signal transduction
XX      system to Ras superfamily cascades. The present sequence
XX      is used for diagnosis and treatment (by protein therapy) of
XX      CAMP-GEF-associated disorders. It is also useful for developing drugs
XX      and producing cell-lines or transgenic animal models for Ras-associated
XX      cancers.
XX
SQ      Sequence 881 AA:

```

```

RESULT 8
AAI70965
ID      AAI70965 standard; Protein; 881 AA.
XX
AC      AAI70965;

```

```

Query Match      8.7%; Score 671; DB 21; Length 881;
Best Local Similarity 23.5%; Pred. No. 2.7e-38;
Matches 229; Conservative 147; Mismatches 287; Indels 312; Gaps 25;

```

[illegible]

AC	AAY70968.
XX	
DT	09-AUG-2000 (first entry)
XX	
DE	Human Ras signalling pathway associated protein CAMP-GEFII.
XX	
KW	Ras signalling pathway; CAMP-GEFII; cyclic adenosine monophosphate;
KW	guanine nucleotide exchange factor; Rap1A; diagnosis; treatment;
KW	CAMP-GEF-associated disorder; drug; transgenic animal model;
KM	Ras-associated cancer; protein therapy; human.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Binding-site
FT	Location/Qualifiers
FT	222..291
FT	/label= "CAMP-binding_domain
FT	/note= "amino acid at position 271 determines binding
FT	specificity for CAMP/cGMP"
FT	606..629
FT	/label= "SCR1
FT	/note= "Structurally conserved region which is
FT	highly homologous to Ras-superfamily GEFs"
FT	678..720
FT	/label= "SCR2
FT	/note= "Structurally conserved region which is
FT	highly homologous to Ras-superfamily GEFs"
FT	757..778
FT	/label= "SCR3
FT	/note= "Structurally conserved region which is
FT	highly homologous to Ras-superfamily GEFs"
XX	
PN	WO200024766-A2.
XX	
PD	04-MAY-2000.
XX	
PF	22-OCT-1999; 99WO-US24826.
XX	
PR	23-OCT-1998; 98US-0105507.
PR	16-NOV-1998; 98US-0108685.
XX	
PA	(MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI	
PI	Kawasaki H, Graybiel A, Hausman D;
XX	
DR	WPI: 2000-350690/30.
DR	N-PSDB: AAD00318.
XX	
PT	Isolated nucleic acid comprises nucleotide sequence encoding protein
PT	selected from normal or mutant CalDAG-guanine nucleotide exchange
PT	factor -
XX	
PS	Claim 64; Page 126-128; 128pp; English.
XX	
XX	The present sequence is a human cyclic adenosine 3', 5'
CC	monophosphate-guanine nucleotide exchange factorII (CAMP-GEFII)
CC	domains. It has substrate specificity for Rap1A and is differentially
CC	distributed in brain and various other tissues. CAMP-GEF proteins
CC	selectively activate the Ras superfamily small G protein and Rap1A,
CC	function as a negative regulator of Ras and directly couple the CAMP
CC	signal transduction system to Ras superfamily cascades. The present
CC	sequence is used for diagnosis and treatment (protein therapy) of
CC	CAMP-GEF-associated disorders. It is also useful for developing drugs
CC	and producing cell-lines or transgenic animal models for Ras-associated
CC	cancers.
XX	
SQ	Sequence 849 AA:
	Query Match 8.6%; Score 664.5; DB 21; Length 849;
	Best Local Similarity 23.8%; Pred. No. 7.3e-38;
	Matches 232; Conservative 152; Mismatches 276; Indels 313; Gaps 32
OY	88 DDDEEDIERASDPLMSDIYRDC-----LEKPIORTDDIE 124

```

Db 129 DDEHEDA-----PLPTEEEKCEDEELQDTMLLSQMGPDAMHMLRKPPGQRTVDLLE 183
QY 125 QLEFPHQOLPAFANMTMSVARELCAYMFAVAVRAGTIYVINDGEELDSNVILINGSVEVT 184
Db 184 IYIEELHAKALSHLSTYKRELAVLIEFSAHKGGLVLENOGEETSMWILKGSNNV 243
QY 185 -YPDGKAELICMNSGFSVPTMDKEYMKGMRTKVDCCQVCAOODYCIILNQVEKNMQ 243
Db 244 IYKGYVCTLHESDDFGKALVNDAPRAASTIVREDNCHFLRYDKEDFNILNDVEANTY 303
QY 244 KVEEGEIVYWKREHRELDRTGRKG-----HIVIKGTSERLTMLVE-----EHSVDP 292
Db 304 RLKEHDQDVLEKVPAGNANSGNSQPOQKXTVMGSGTEKLEHLEHLEATLNEA 363
QY 293 T--FIEDFLITTYFTLSSPMEVGKLLMFNDLSL-----RDKVTRVVLLMY 337
Db 364 TDSVLNDFIMMHCVFPMFTNQLCPALVAHYHAOPSGOGEKMDYALNNKRRVRLVLOWA 423
QY 338 NNHFNDEGDPAMTRFLEEFENLREKMGHILRLINACAARRLMTLTK-----PSR 393
Db 424 AMYGDLQEDDVSMALFEETVYSVD-----ARMT-----AALKQDLPLEKIVKQISE 473
QY 394 EAPLPFILLGSEKGFIFVDSVSGSKATE--AGLKRGDQILEVNGQNFENIQLSKAME 451
Db 474 DAKAP-----QKHKVLLQENTGDERAQKQPIRGSEV-----LKVYVC 514
QY 452 ILRNNHLSITVYTNLFEKELLRLSEKRNQAPHLPIKIGIKASRSIPDLAVDEQ 511
Db 515 MDHTYTTIRPVATSV---KEVISAADKLGSGE-----545
QY 512 VIGLEKYNKSKANTYGGRNKILDKTRISLPOKPYNDIGIGSQDSDIYGLQTKH 571
Db 546 -GLIIVKMS-----GGE---KVL-----KP-NDVSVF-----569
QY 572 IPTALPVSGTLSSNPDLQSHHRIIDFSATPDLPDQVLEFRADQSRYIMISKDTAK 631
Db 570 -TTLTNG-----RLF-----579
QY 632 EAVYQAIREFAVATATP-DQYSLCEVSYTPREGVIKORRLPDQSLKADRIOLSGRYIYKNN 650
Db 580 -----ACPREQPD-----SLTP-----LPEQEGPTVGTV---GTFEL---608
QY 691 METETLCSDEDAQELLRESQISLLQSTVEATQLSMRNELFRNIPEPTYIDLF-KLR 749
Db 609 -----MSSKDLAYQMTTYDMLFCVHELITLHTFERHN 643
QY 750 SKTSCANLKRFEVINOETFWASSELRETNOILKRMKIIHKFIKIALHCREKNFNSMFA 809
Db 644 FKTTANLDFLRFRNIOFWVTEICLSQLSKRQOLKFKIKIAHCKREYNLNSFEA 703
QY 810 IISGLNLAPYARLKTWEKLPNKYEKLEFDLQDLFDBSRMAKRYRVNLSONTLOPPIPL 869
Db 704 IYVGLSIAVSRALITWEKLPSEKFKFYAEFESIMDSRRHRAVR--LTVAKLEPLIIF 761
QY 870 FPIYIKDQFLHGNOSKVGVLNFEKLMIAKEIRIVGMAVYNNDPALMFTRRKKWR 929
Db 762 MPLLIKDMTTHGKNKFTIDNLNFEKMRMIANTARTVYRSOPFNP-----809
QY 930 SLGSLSGSSTNATVLDYAQTGCHKKRVRSSFLNAKKLYEDAOAMARKVOYLNSLELEMD 989
Db 810 -----DAAQ-----ANKNHQ-----VRSYVQLANVIDN 833
QY 990 EESLQTLISLOCEP 1002
Db 834 QRTLSOMSHRLEP 846

```

RESULT 10  
 ID ABB59888 standard; Protein; 957 AA.  
 AC ABB59888;

```

XX XX
DT 26-MAR-2002 (first entry)
XX XX
DE Drosophila melanogaster polypeptide SEQ ID NO 6456.
XX XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX XX
OS Drosophila melanogaster.
XX XX
PN WO200171042-A2.
XX XX
PD 27-SEP-2001.
XX XX
PF 23-MAR-2001; 2001WO-US09231.
XX XX
PR 23-MAR-2000; 2000US-191637P.
XX XX
PR 11-JUL-2000; 2000US-0614150.
XX XX
PA (PEKE ) PE CORP NY.
XX XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX XX
DR MPI; 2001-656860/75.
XX XX
DR N-PSDB; ABL03991.
XX XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX XX
PS Disclosure: SEQ ID NO 6456; 21pp + Sequence Listing; English.
XX XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX XX
SQ Sequence 957 AA:

```

Query Match 8.3%; Score 637; DB 22; Length 957;  
 Best local Similarity 23.8%; Pred. No. 7.8e-36;  
 Matches 212; Conservative 156; Mismatches 272; Indels 252; Gaps 29;

```

QY 62 IYATESEAGMDMLSGLPETAVDSED-DDEEDIERASDPLMSRD---IVRDCLEKDPID 117
Db 244 LYFRIDIEEGSTAAGVPGQ-----AEDIGANEHIREALSALFORGPATRLMLRKPSHE 299
QY 118 RTDDDEIQLLEFPHQOLPAFANMTMSVARELCAYMFAVAVRAGTIYVINDGEELDSNVIL 177
Db 300 RTSEELVEFEELVHAALSHLSTSIKRELSSIFVEAHQAQCTIIFNOGDESRWYIL 359
QY 178 NGSVEVT-YPDGKAELICMNSGFSVPTMDKEYMKGMRTKVDCCQVCAOODYCIILN 236
Db 360 KGSVDVYIHGKGYATLTKGDDFGKALVNDAPRAASTIVKKNCHLRRDKHEFRILR 419
QY 237 QVERKNQVEEGEIVYWKREHRELDRTGRKG-----HIVIKGTSERLTMLVEE--H 287
Db 420 DVEANTLRLQEHGKDVLY-----LERYAKORGQHSAFKYTVMGTFPAKMLEHLELTRIGQ 474
QY 288 SV--VDPFTIEDFLITTYFTLSSPMEVGKLLMFNDPSLBDKVTRVVLLMYNN 339
Db 475 SVGMDP-FIUDFLTLHYMFAVYVQDLDAQTPEDEREYIT-----NEKKRVIOFMQKVM 527
QY 340 HFNDDEGDPAMTRFLEEFENLREKMGHILRLINACAARRLMTLTKPSREAPLPF 399
Db 528 VRHAAPFEPSVCFIEDLAAVEADP-----553

```

```

QY 400 ILGSGEKGCEIVDSVDSGSKATEAGLKGDDILEVNGONFENIOLSKAMEILRNTHL 459
D 554 -----DLNEBTSTVHN----- 564
QY 460 SITVKNLFEVKEKELLTRLS--EKKRNGAPHLPIGDIKASRSYSPDLAVDEQVIGLEK 517
D 565 -----VLTMARQEDRN-----QN 579
QY 518 VNKSANTYGGRNKLKILDKTRISILPKOPYNDIGSGSDSIVGLRQTKHIPALP 577
D 580 AGQWK-----LP-----PNCOP 592
QY 578 V---SGTSSNPDLLOSHRIIDFSAITPDLPOVLVFAKADQOSRYIMSKDTAKEV 634
D 593 ICLFSGNATPSKTYI-----RPD-DDILFRYCCADHTCYTLRPPMHTAELIK 639
QY 635 IQATREFAVTPTPOYSICEVSTPE-GVIKOR--RLPDOLSKLADRIOLSGRY--LKN 689
D 640 ACADKRLQNLNGPDLVLYEVKNGERSVFEDNDVSIPTGIS-----LNGRLFVSKD 692
QY 690 NMETETLCSDDAQELLRES-QISLLOLSTIVEVATQLSMRNEELFRNIETPTVIDLF-K 747
D 693 HLDALTOQE---OCEPTEGVIDLEILSTKELAVHITLFEMDLFWAVEHELYHTEGR 749
QY 748 LRSKTSANLKRFEVINQETFWVASLRETNOLKRMKIIKHFIKIALHORECKNNSM 807
D 750 HHFGKITANLQVFLRRFNEVOYVITELVSTPSLSKRGVLKRAIKLAAYCKEYONLMAF 809
QY 808 FAISGLNLAFVARTLTWETKLPNKYEKLFODLQDLFDPSSHMAKYRNVLNSQNLQPPII 867
D 810 FAVVGLSNMVAVSRLQGWKEXIKPSFKRIQEFELIDPSRNHRAYRFGV--KLQPLLI 867
QY 868 PLFPIYTKDILFLHGGNSKVDGLVNFELKMLAKETIHVG--KMASVNMMP 917
D 868 PEMPLLLDMTFEAFHGNKTSLDGLVNFEMKMMMAQMTMTIRFCRSRSLGEP 919

RESULT 11
ID AAB07793 standard; Protein: 262 AA.
AC AAB07793;
DT 07-NOV-2000 (first entry)
DE Partial murine guanine nucleotide releasing factor 4.
XX
KW guanine nucleotide releasing factor 4; GRF-4; Ras activator;
KW Nedd4 ubiquitination; cell metabolism; cell proliferation; cancer;
KW cell differentiation; cell transformation; neuronal disorder.
XX
OS Mus sp.
PN MO200043510-A2.
PD 27-JUL-2000.
PF 20-JAN-2000; 2000WO-CA00042.
PR 20-JAN-1999; 99CA-2259830.
PA (HSCR-) HSC RES & DEV LP.
XX
PI Rottin D, Pham N;
XX WPI; 2000-499228/4.
DR N-PSDB; AAA59384.
XX
PT Nucleic acids encoding guanine nucleotide releasing factor-4 useful for
PS the treatment of cancers and neuronal disorders -
Claim 34; Fig 19b; 89pp; English.

```

```

XX
CC AAB07793-95 represent partial murine guanine nucleotide releasing
CC factor (GRF)-4 (Ras activator) polypeptides; encoded by the 3
CC reading frames of clone 7.7 (AAA59384). GRF4 activates Ras both
CC in vitro and in vivo. It directly binds cyclic adenosine monophosphate
CC (cAMP) directly via its CNMP-BD (cAMP/guanine monophosphate (cGMP)
CC binding domain). GRF4 directly connects cAMP-generating (e.g. G protein
CC coupled receptors) or cGMP-generating pathways to Ras. GRF4 activates
CC Ras in response to elevation of intracellular cAMP and/or cGMP. GRF4
CC is a target for Nedd4 ubiquitination as it binds Nedd4. Activation of
CC the Ras signalling pathway controls numerous cellular functions, such as
CC cell metabolism, proliferation, differentiation and transformation.
CC Therefore modulation of Ras activity may provide a mechanism for
CC controlling diseases. GRF4 polynucleotides and polypeptides may be used
CC in the treatment of diseases associated with inappropriate GRF4
CC expression and activity such as cancers and neuronal disorders. The
CC GRF4 polypeptides may be used as antigens in the production of
CC antibodies against GRF4 and in assays to identify modulators (agonists
CC and antagonists) of GRF4 expression and activity. The anti-GRF4
CC antibodies and GRF4 antagonists may also be used to down regulate GRF4
CC expression and activity. Inhibition of Ras can reduce cellulose
CC proliferation and cancers.
CC
XX
SQ Sequence 262 AA;
Query Match 7.1%; Score 545.5; DB 21; Length 262;
Best Local Similarity 66.9%; Pred. No. 2.9e-30;
Matches 107; Conservative 13; Mismatches 37; Indels 3; Gaps 2;

QY 1325 SWASSTGYWGEDSEGDGTITKRRGKDVSIIEASSLTVTTEIKPVMPAHIAVAST 1384
D 6 SMSSTAAVAALLEVLDPPCGCHNSGCKDVSAEASSSMVPTTEAKPVMPAHIAVTPST 65
QY 1385 TKGLIAKRGYREPPPTPGYIGITPTPEEGHSPARKPPYVALORSRVARSST 1444
D 66 TKGLIAKRGYREPPPTPGYIGITPTPEEGHSPARKPPYVALORSRVARSST 125
QY 1445 AGPSSVQOPGHPTSSRPVKNPQMHKRNESDPRLAARYQSO 1484
D 126 PABG--QTPPA-AAASRSGSKPMWHKPSDADPRLARAFOPQ 162

RESULT 12
ID AAM59534 standard; Protein: 119 AA.
AC AAM59534;
DT 05-NOV-2001 (first entry)
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 31639.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
PN MO200157275-A2.
PD 09-AUG-2001.
PF 30-JAN-2001; 2001WO-US00667.
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.

```

```

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI: 2001-483446/52.
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains -
XX PS Example 4; SEQ ID NO: 31639; 650pp + Sequence listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC brain. They can be used to measure gene expression in brain cell samples,
XX CC which may enable the diagnosis and improved treatment of nervous system
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX CC epilepsy and cancers. The present sequence is a protein encoded by one of
XX CC the probes of the invention.
XX SQ Sequence 119 AA;

Query Match 7.0%; Score 543; DB 22; Length 119;
Best Local Similarity 85.7%; Pred. No. 1.2e-30;
Matches 102; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 685 YLKNMTEETLCSDDAQELRESQISLQLSTVEATQLSMRNELEPNIEPTETID 744
DB 1 YLKNMTEETLCSDDAQELRESQISLQLSTVEATQLSMRNELEPNIEPTETID 60

QY 745 LFLRSKTSKANLKEEVINQETFWVASEILRETNOLKMKIKIHKFIKIALHCRECKN 803
DB 61 LFLRSKTSKANLKEEVINQETFWVASEILRETNOLKMKIKIHKFIKIALHCRECKN 119

RESULT 13
AAM72102
ID AAM72102 standard; Protein: 119 AA.
XX AC AAM72102;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 32408.
XX KW Human: bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX PT WPI: 2001-488900/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow -
XX Example 4; SEQ ID NO: 32408; 658bp + Sequence listing; English.

```

```

CC CC The present invention provides a number of single exon nucleic acid
CC CC probes which are derived from genomic sequences expressed in the human
CC CC bone marrow. They can be used to measure gene expression in bone marrow
CC CC samples, which may enable the improved diagnosis and treatment of cancers
CC CC such as lymphoma, leukemia and myeloma. The present sequence is a
CC CC protein encoded by one of the probes of the invention.
XX SQ Sequence 119 AA;

Query Match 7.0%; Score 543; DB 22; Length 119;
Best Local Similarity 85.7%; Pred. No. 1.2e-30;
Matches 102; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 685 YLKNMTEETLCSDDAQELRESQISLQLSTVEATQLSMRNELEPNIEPTETID 744
DB 1 YLKNMTEETLCSDDAQELRESQISLQLSTVEATQLSMRNELEPNIEPTETID 60

QY 745 LFLRSKTSKANLKEEVINQETFWVASEILRETNOLKMKIKIHKFIKIALHCRECKN 803
DB 61 LFLRSKTSKANLKEEVINQETFWVASEILRETNOLKMKIKIHKFIKIALHCRECKN 119

RESULT 14
AAM32364
ID AAM32364 standard; Protein: 119 AA.
XX AC AAM32364;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #6401 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX PT WPI: 2001-488897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
XX PS Claim 27; SEQ ID NO 32633; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP:
XX CC see AAI31315-AA157546). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for producing a microarray for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from human placenta. The probes are useful for antenatal diagnosis of
XX CC human genetic disorders.
XX SQ Sequence 119 AA;

Query Match 7.0%; Score 543; DB 22; Length 119;
Best Local Similarity 85.7%; Pred. No. 1.2e-30;

```





REFERENCE AUTHORS	3 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuwa, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
TITLE	Riken integrated sequence analysis (RISA) system-384-format
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE PUBMED REFERENCES	20530913 11076861
AUTHORS	4 Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Atakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, M., Gaasterland, T., Glisic, C., Kling, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikola, I., Pesole, G., Quackenbush, J., Schraml, L.M., Straub, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barish, G., Blake, J., Bonfield, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.T., Bulc, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hotman, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombarts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Togo, Oka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S., and Hayashizaki, Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
MEDLINE PUBMED REFERENCES	21085660 11217851
AUTHORS	5 (bases 1 to 163) Adachi, J., Aizawa, K., Akahira, S., Akinura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bulc, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayasu, N., Hill, D., Hiramoto, K., Hirooka, T., Hoti, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Koude, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, K., Sakai, K., Sano, H., Sasaki, D., Schraml, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Soabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Teijima, Y., Toya, T., Yamamura, T., Yamataka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel.: 81-45-503-9222. Fax: 81-45-503-9216)
COMMENT	Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details. CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAAGAGACCAATCCAGAGCTCTTTTGTGGTTTTTNN 3']. cDNA was prepared by using reverse transcriptase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot -10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5'-



```
Db 647 TTGTAATTTAGGGTTTGAAGCTTTTGTATTTATTTATTAAGAAACATTATTAAGT 706
QY 5607 GGATACATTTGCAAGTAAGACAGCTTGGATGTGTGAGCTAATGCCAGCTTTATACAG 5666
Db 707 GGATACATTTGCAAGTAAGACAGCTTGGATGTGTGAGCTAATGCCAGCTTTATACAG 766
QY 5667 CTTCTTCAAGACAGCTTCCCTTATTTGAATTTGGCATTTAGGGAATTAACAGCTTTAAAC 5726
Db 767 CTTCTTCAAGACAGCTTCCCTTATTTGAATTTGGCATTTAGGGAATTTAAACAGCTTTAAAC 817
QY 5727 GTGATTAAGATCAAAACCTGGTATGACATGCCAGCTTTTGCAGAGGAGTTAGTACAC 5786
Db 818 ATGATTAAGATCAAAACCTGGTATGACATGCCAGCTTTTGCAGAGGAGTTAGTACAC 872
QY 5787 AAGACTAATCTCCAGTGGCTTTATGAGCGCTGCATATAGAGAGGCTTAAGTACGA 5846
Db 873 AAGGCTGACCTGTAAGTGGCTTTGGGAGCGCTGATGAGAGAGGCC-AAGTGTAGCG 931
QY 5847 ACCATCTGCTACAGCTGCTATTAACCTATTAATGACTG-AAATGACCCCTCCACTCTAT 5905
Db 932 CCGGTCTGCTACAGCTGCTGCTAGCCCTTAGATGACTGAAATGACCCCGGCTCTAT 991
QY 5906 TTTTGTGTG-TTTTGCACAGACTCCGGAAGAGTGAAGGCTGCCAATCTGAGTAGTACT 5964
Db 992 TCTTGTGTGTTTTCACAGACTCCGGAAGAGTGAAGGCTGCCAATCCGAGTAGTACT 1051
QY 5965 AATGTGAGAGTCTGCTGCTGCTGATTTTTCATTAAATTCAGCTGATCATATGA 6024
Db 1052 AGATGTGAGGCTGCTGCTGCTGCTGGA--TTTTCATTAAATTCAGCTGATCATATGA 1108
QY 6025 TCAGTAGATTAAGCTAAATAGCTTCAAAATTTTAAAGTGAATTGAGCTTTTTCAC 6084
Db 1109 TCGATGATTAAGCTAAATAGCTTCAAAATTTTAAAGTGAATTGAGCTTTTTCAC 1168
QY 6085 GATTCAAACAGTCTAGCTGCTTATTTATTAATTAATCTCTCTGATGATGAGCTATGCTA 6144
Db 1169 GATTCAAACAGTCTAGCTGCTTATTTATTAATTAATCTCTCTGATGATGAGCTATGCTA 1228
QY 6145 CTGCTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6204
Db 1229 CTGCTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1287
QY 6205 CAGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6264
Db 1288 CAGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1347
QY 6265 ACCACAAGTTGCACAAATGTTATCTAAGCATTAAGTAATTTGAGAAATAGAGACTGCTA 6324
Db 1348 ACCACAAGTTGCACAAATGTTATCTAAGCATTAAGTAATTTGAGAAATAGAGACTGCTA 1407
QY 6325 TCTCAGTTGCTCTGTGATGTCAAGTGCAGATGTACAAATTAACCTGCTGCTCATA 6384
Db 1408 TCTCAGTTGCTCTGTGATGTCAAGTGCAGAAATGTACAAATTAACCTGCTGCTCATA 1467
QY 6385 CTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6444
Db 1468 CTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1527
QY 6445 TTTTGAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTTTGTGAAGATGCTA 6504
Db 1528 TTTTGAATTTTAAATTAATTAATTAATTAATTAATTAATTTTGTGAAGATGCTA 1587
QY 6505 AATGTACTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6564
Db 1588 AATGTACTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1647
QY 6565 ACAT 6568
Db 1648 ACAT 1651
```

RESULT 2  
AL583554

```
LOCUS AL583554 789 bp mRNA linear EST 16-FEB-2001
DEFINITION AL583554 LFL1_FLO13_FBRn1 Homo sapiens cDNA clone CS0DF006YF09 5
prime, mRNA sequence.
ACCESSION AL583554
VERSION AL583554.1 GI:12952629
KEYWORDS
SOURCE
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 789)
AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..789
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF006YF09"
/clone_1ib="LFL1_FLO13_FBRn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a Noli-oligo(dt) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Reng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
http://fulllength.invitrogen.com"
BASE COUNT 247 a 130 c 134 g 273 t 5 others
ORIGIN
Query Match 11.6%; Score 760.8; DB 9; Length 789;
Best Local Similarity 99.3%; Pred. No. 1,8e-143;
Matches 759; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 5805 GGCCTTATGAGCCCTGCATATAGAGAGCCCTAAGTGTAGCAACCATCTGCTACAGCTG 5864
Db 1 GGCCTTATGAGCCCTGCATATAGAGAGCCCTAAGTGTAGCAACCATCTGCTACAGCTG 60
QY 5865 CTAATTAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5924
Db 61 CTAATTAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 120
QY 5925 GACTCGGAAGATGAAGGCTGCCAATCTGAGTAGTACTCAAAATGAGAGAGCTGCTG 5984
Db 121 GACTCGGAAGATGAAGGCTGCCAATCTGAGTAGTACTCAAAATGAGAGAGCTGCTG 180
QY 5985 CTTGGAATTTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6044
Db 181 CTTGGAATTTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 240
QY 6045 GCTTCAATTTTAAAGTGAATTTGAGAGTCTTTTTCACGTATCAAAACAAATGCTAGTC 6104
Db 241 GCTTCAATTTTAAAGTGAATTTGAGAGTCTTTTTCACGTATCAAAACAAATGCTAGTC 300
QY 6105 TTTATTTAATTAATTTCTCTGTATGATGATGATGATGATGATGATGATGATGATGATG 6164
Db 301 TTTATTTAATTAATTTCTCTGTATGATGATGATGATGATGATGATGATGATGATGATG 360
QY 6165 TTTATGATTTTGTATTTTATGATGATGATGATGATGATGATGATGATGATGATGATG 6224
Db 361 TTTATGATTTTGTATTTTATGATGATGATGATGATGATGATGATGATGATGATGATG 420
QY 6225 TGGAGCTCATGTCATATATAGAAAGACGAATATGCTGCTACCAAGTTGCAAAATGT 6284
```

RESULT 3	AU140423	769 bp	mRNA	linear	EST 05-AUG-20022
LOCUS	AU140423				
DEFINITION	AU140423	PLACE3	Homo sapiens	CDNA clone	PLACE300032 5', mRNA
ACCESSION	AU140423				
VERSION	AU140423.1	GI:11001944			
KEYWORDS	EST.				
SOURCE	human.				

Helix Research Institute  
1532-3 Yama, Kisarazu, Chiba 292-0812, Japan  
Tel.: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: [genomies@rti.co.jp](mailto:genomies@rti.co.jp)  
HRI Human CDNA project, 5'- and 3'-end one pass sequencing: Helix  
Research Institute: cDNA library construction: Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

Query Match	11.2%	Score 736.2	DB 9	Length 769
Best Local Similarity	98.7%	Pred. No. 1.7e-138		
Matches 760; Conservative	0	Mismatches 8	Indels 2	Gaps 2

Oy 2491 TCATCAGTGGCCCTAAACCTGSCACCACTGCGCAACGACCTGGGAGAAATTC 2550  
|||||  
Db 1 TCATCAGTGGCCCTAAACCTGSCACCACTGCGCAACGACCTGGGAGAAATTC 60

QY	2851	GTTTGGGCTCTCTCAGCCAGGGTGTCAAAATGCACAGTGTAGTATTTGGTCTCAGACG	2910
Db	361	GTTTGGGGCTCTCAGCCAGGGTGTCAAAATGCACAGTGTAGTATTTGGTCTCAGACG	420
QY	2911	GTGGTCATAAAAAGCGGGTA-CGTGTAAGTTCCTTCTCAATGCCAAAAAGCTTATGAA	2966
Db	421	GTGGTCATAAAAAGCGGGTACCGTGTAGTTCCTTCTCAATGCCAAAAAGCTTATGAA	480
QY	2970	GATGCCCAATGGCTCGAAAAGTAAAGACAGTACTTTCCAATTGGAGCTAGAAATGAGC	3026
Db	481	GATGCCCAATGGCTCGAAAAGTAAAGCGGTACTTTCCAATTGGAGCTAGAAATGAGC	540
QY	3030	GAGGAGAGCTTCAGACATTATCTGTGAGTGAGCGACGACACACACATTGGCTAAG	3088

[illegible]

RESULT 4					
AL534790/c					
LOCUS	AL534790	743 bp	mRNA	linear	EST 13-FEB-2001
DEFINITION	AL534790 L1T_FL013_Fbrn1	Homo sapiens	CDNA	clone CS0DF006YF09 3	
	prime, mRNA	sequence.			
ACCESSION	AL534790				
VERSION	AL534790.1	GI:12798283			
KEYWORDS	EST.				
SOURCE	human.				

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 743)  
REFERENCE  
AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
CONTACT Contact: Genoscope  
COMMENT Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: [genoscope.cns.fr](mailto:genoscope.cns.fr) Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)

FEATURES	Location/Qualifiers
source	1. .743 /organism="Homo sapiens" /db_xref="taxon:9606"

```

/clone="CSDF006FE09"
/clone.lib="LTI.FL013.FBrn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
/notes="Organ: Fetal brain; Vector: PCWSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
PCWSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
Maryland 20850, USA Fax : (1) 301 610 8371 Email :
filiang@lifestech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT      271 a      132 c      125 g      209 t      6 others
ORIGIN

Query Match      11.1%; Score 727.4; DB 9; Length 743;
Best Local Similarity 99.1%; Pred. No. 1e-136;
Matches 736; Conservative 5; Mismatches 1; Indels 1; Gaps 1;

OY 5805 GGCCTTATGAGCCCTGCATATAGAGAGCCCTAAGTGTAGCAACATCTGCTCACAGCTG 5864
|||||
DB 743 GGCCTTATGAGCCCTGCATATAGAGAGCCCTAAGTGTAGCAACATCTGCTCACAGCTG 684

OY 5865 CTAATTAACCTTAATAGATGAGAAATGACCCCTCAGCTATTTTGTGTGTTGACACA 5924
|||||
DB 683 CTAATTAACCTTAATAGATGAGAAATGACCCCTCAGCTATTTTGTGTGTTGACACA 624

OY 5925 GACTCCGGAAGAGGAGGCTGCAATCTGAGTAGTACTCAATGAGAGAGCTGCTGT 5984
|||||
DB 623 GACTCCGGAAGAGGAGGCTGCAATCTGAGTAGTACTCAATGAGAGAGCTGCTGT 564

OY 5985 CTTCGATTTTTCCTATTAATTCAGCTGATATATGATGCTGATATTAACGTAATA 6044
|||||
DB 563 CTTCGATTTTTCCTATTAATTCAGCTGATATATGATGCTGATATTAACGTAATA 504

OY 6045 GCTTCAATTTTAAAGTGAATGAGATGTTTTCACAGTATCAACAAAGTCAAGTGC 6104
|||||
DB 503 GCTTCAATTTTAAAGTGAATGAGATGTTTTCACAGTATCAACAAAGTCAAGTGC 444

OY 6105 TTTATTTAATTAATTCCTCTGATATGAGATGCTATTTGCTATTTATTAATTTGCA 6164
|||||
DB 443 TTTATTTAATTAATTCCTCTGATATGAGATGCTATTTGCTATTTATTAATTTGCA 384

OY 6165 TTTATGATTTGTAATTTTACATGTAATGCAATTTTGGCAGTTTATTAATTAAGCTA 6224
|||||
DB 383 TTTATGATTTGTAATTTTACATGTAATGCAATTTTGGCAGTTTATTAATTAAGCTA 324

OY 6225 TGGACCTCATGTCATATAGAGAGCAAGAAATGCTGCTACCAAGAGTGCACAAATGT 6284
|||||
DB 323 TGGACCTCATGTCATATAGAGAGCAAGAAATGCTGCTACCAAGAGTGCACAAATGT 264

OY 6285 TATCTAAGCATTAAGTAATTTAGAACATAGAGCTGCTATTAATCTCAGTCTGCTGATG 6344
|||||
DB 263 TATCTAAGCATTAAGTAATTTAGAACATAGAGCTGCTATTAATCTCAGTCTGCTGATG 204

OY 6345 TCAAGTGCAGAAATGTACAAATTAACGTGATTTCCCTCATATCTTTGATTAATCTTGACC 6404
|||||
DB 203 TCAAGTGCAGAAATGTACAAATTAACGTGATTTCCCTCATATCTTTGATTAATCTTGACC 144

OY 6405 TGAATGCTTTTGAAGAGACATGAGAGTGTATCCCTTTGTAATTTTATTAACAAT 6464
|||||
DB 143 TGAATGCTTTTGAAGAGACATGAGAGTGTATCCCTTTGTAATTTTATTAACAAT 84

OY 6465 AATTTGACATATGTTATTTTGTGTAAGATGTAGAAATGTAC-TATGTTTATGCT 6523
|||||
DB 83 AATTTGACATATGTTATTTTGTGTAAGATGTAGAAATGTACNTATGTTTATGCT 24

OY 6524 TCTACATCCAGTTTGTACAAAGCT 6546
|||||
DB 23 TCTACATCCAGTTTGTACAAAGCT 1

```

```

RESULT 5
BG403618
LOCUS      BG403618      812 bp      mRNA      linear      EST 12-MAR-2001
DEFINITION 602419334F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4526168 5',
mRNA sequence.
ACCESSION  BG403618
VERSION    BG403618.1  GI:13297066
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 812)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.jnl.gov
            Plate: LHAM10433 row: d column: 09
            High quality sequence stop: 687.
            Location/Qualifiers
                1. 812
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_image="4526168"
                    /clone_lib="NIH_MGC_93"
                    /issue_type="translational cell papilloma, cell line"
                    /lab_host="DH10B (phage-resistant)"
                    /notes="Organ: bladder; Vector: PCMV-SPORT6; Site:1: NotI;
                    Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
                    Average insert size 1.7 kb. Library enriched for
                    full-length clones and constructed by Life Technologies.
                    Note: this is a NIH-MGC Library."

BASE COUNT      248 a      139 c      153 g      272 t
ORIGIN

```

```

Query Match      10.8%; Score 708.8; DB 12; Length 812;
Best Local Similarity 95.9%; Pred. No. 5.9e-133;
Matches 771; Conservative 0; Mismatches 27; Indels 6; Gaps 4;

OY 5758 GCGAGCCTTGCAGAGGTTGATGTCACCAAGAGTCAAGTCCAGAGGCTTATGACG 5817
|||||
DB 1 GCGAGCCTTGCAGAGGTTGATGTCACCAAGAGTCAAGTCCAGAGGCTTATGACG 60

OY 5818 CTGCATATAGAGAGGCTTAAGTGTAGCAACATCTGCTCACAGCTGCTTTAACTAT 5877
|||||
DB 61 CTGCATATAGAGAGGCTTAAGTGTAGCAACATCTGCTCACAGCTGCTTTAACTAT 120

OY 5878 AATGACTGAATGACCCCTCACTATTTTGTGTTGTTGACAGACTCCGAAAAG 5937
|||||
DB 121 AATGACTGAATGACCCCTCACTATTTTGTGTTGTTGACAGACTCCGAAAAG 180

OY 5938 TGAAGGCTGCATCTGAGTAGTACCAATGTGAGGAAGTCTGCTGGAATTTT 5997
|||||
DB 181 TGAAGGCTGCATCTGAGTAGTACCAATGTGAGGAAGTCTGCTGGAATTTT 240

OY 5998 TCCATTAAATTCAGCTGATATTTGATCATAGATAAAGCTAAATAGCTTCAAAATTTTA 6057
|||||
DB 241 TCCATTAAATTCAGCTGATATTTGATCATAGATAAAGCTAAATAGCTTCAAAATTTTA 300

OY 6058 AAAGTGAATTCAGTGTTTTTCACGTATTCATCAACAAATGCTGCTTATTAATTAAT 6117
|||||
DB 301 AAAGTGAATTCAGTGTTTTTCACGTATTCATCAACAAATGCTGCTTATTAATTAAT 360

```

QY	6118	TCTCTTCTGATCATNGCATTTCCTCACTTCGCTATATACATGTCATTAATGATGATTTGTA	6127		
Db	361	TCTCTTCTGATCATNGCATTTCCTCACTTCGCTATATATACATGTCATTAATGATGATTTGTA	420		
QY	6178	ATTTTACATGTAAATATGCAATATATTTGGCAGTTTATATATAGCTATAGACCTCATGTG	6237		
Db	421	ATTTTACATGTAAATATGCAATATATTTGGCAGTTTATATATAGCTATAGACCTCATGTG	480		
QY	6238	CATATAGAAAGACAGAAATCTAGCTCTACACACAGTGCACAAATGTTATGTAAGCATTA	6297		
Db	481	CATATAGAAAGACAGAAATCTAGCTCTACACACAGTGCACAAATGTTATGTAAGCATTA	540		
QY	6298	ACTTAATTTGAGAAACATPAGGACTGCTAAATCTGCTGCGCTGCTGATGTCAAGTCAGAAAT	6357		
Db	541	AGTAATTTGAGAAACATPAGGACTGCTAAATCTGCTGCGCTGCTGATGTCAAGTCAGAAAT	600		
QY	6358	GTACATATTACTGTGATTTCCCTCATCTACTTTTGATATGATGATGATGATGCTTTTGA	6417		
Db	601	GTACATATTACTGTGATTTCCCTCATCTACTTTTGATATGATGATGATGATGCTTTTGA	659		
QY	6418	GAAAGCATTTGTTGGAGTGTGATCCCTTTTGTATATTTTAATACAAATATGTCATATT	6477		
Db	660	GAAAGCATTTGTTGGAGTGTGATCCCTTTTGTATATTTTAATACAAATATGTCATATT	718		
QY	6478	GGTTATATTTTGTTCGAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG	6537		
Db	719	GG---TATATTAAGTGAAGATGTGTGAAGATGATGATGATGATGATGATGATGATGATG	774		
QY	6538	GTACAAAGCTGGAATAATTAATTAAT	6561		
Db	775	GGAAAGCTGGAATAATTAATTAAT	798		
RESULT 6	BO017652/c	732 bp	mRNA	linear	EST 27-MAR-2002
LOCUS	UI-H-EDD-awy-b-01-0-UI-s1 NCI-CGAP_ED0				Homo sapiens cDNA clone
DEFINITION	IMAGE:5825184 3', mRNA sequence.				
ACCESSION	BO017652				
VERSION	BO017652.1	GI:19752929			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
COMMENT	Unpublished (1997)				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgaabs-r@mail.nih.gov				
	Tissue Procurement: Dr. Jose Mercuende				
	cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa				
	DNA Sequencing by: Dr. M. Bento Soares, University of Iowa				
	Clone Distribution: Clone distribution information can be found				
	through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov				
	Seq primer: M13 FORWARD				
	PolyA=yes.				
FEATURES	Location/Qualifiers				
source	1..732				

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5825184"
/clone_lib="NCI CGAP_E00"
/tissue_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH03 (Life Technologies)"
(note="Organ: Left Pubic Bone; Vector: pT73-Pac
(Pharmacia) with a modified polylinker; Site:1; Ecor I;
Site 2: Not I; NCI CGAP_E00 is a cDNA library containing
the following tissue(s): Chondrosarcoma cell line C55. The

```

library was constructed according to Bonaldó, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)<sub>18</sub> tail. The sequence tag for this library is GCTCAAGGCT.

TAG\_L1B-UI-H-ED0  
TAG\_TISSUE=chondrosarcoma  
TAG\_SEQ=CGTCAAGGCT"

BASE COUNT	267 a	120 c	114 g	231 t
ORIGIN				
Query Match	10.5%;	Score 691;	DB 14;	Length 732;
Best Local Similarity	99.9%;	Pred. No. 2.4e-125;		
Matches 702; Conservative	0;	Mismatches 0;	Indels 1;	Gaps 1

Qy	5866	TATTAACCTTATATAGCTGAATAGAACCCCTCAGCTCTATTTTGGTGTGTTTCACAG	5925
Db	732	TATTTAACCTTATATAGCTGAATAGAACCCCTCAGCTCTATTTTGGTGTGTTTCACAG	673
Qy	5926	ACTCGGAAAAGTGAAGGCTGCCAATCTGAGTAGTACTGAAATGTGAGGAGCTGCTGTC	5985
Db	672	ACTCGGAAAAGTGAAGGCTGCCAATCTGAGTAGTACTGAAATGTGAGGAGCTGCTGTC	613
Qy	5986	TTTGCATTTTTTTCCTTAATATTCAGCTGATTCATATGATGACGTGATTAAGCTAAATAG	6045
Db	612	TTTGGATTTTTTTCCTTAATATTCAGCTGATTCATATGATGACGTGATTAAGCTAAATAG	554
Qy	6046	CTTCAAAATTTTAAAAGTGAAGTTCGACATGTTTTTTTACAGTATCAAAACAATGTCAGTCT	6105
Db	553	CTTCAAAATTTTAAAAGTGAAGTTCGACATGTTTTTTTACAGTATCAAAACAATGTCAGTCT	494
Qy	6106	TTATTTAATAATTCCTCTCTGATCATGACATTTTCTACTGCTTATTAATTCATATTCATAT	6165
Db	493	TTATTTAATAATTCCTCTCTGATCATGACATTTTCTACTGCTTATTAATTCATATTCATAT	434
Qy	6166	TATGCAATTTGTAATTTTATTCATGTAATATGCAATATTTTGGCAGTTTATTTATATAGGCTAT	6225
Db	433	TATGCAATTTGTAATTTTATTCATGTAATATGCAATATTTTGGCAGTTTATTTATATAGGCTAT	374
Qy	6226	GGACCTCATGTGCATATATGAAAAGACAGAAATCTTAGCTTACCAACAAGTTGCACAAATGTT	6285
Db	373	GGACCTCATGTGCATATATGAAAAGACAGAAATCTTAGCTTACCAACAAGTTGCACAAATGTT	314
Qy	6286	ATCTAGACATTAAGTAATTTGATAGAACTATAGACATGCTAATCTAGTTCGCTCTGATGATG	6345
Db	313	ATCTAGACATTAAGTAATTTGATAGAACTATAGACATGCTAATCTAGTTCGCTCTGATGATG	254
Qy	6346	CAAGTGCAGAAATGTAACAATTAAGTGGTGTATTTCTCATACTTTGATTAATCTTGTAGCT	6405
Db	253	CAAGTGCAGAAATGTAACAATTAAGTGGTGTATTTCTCATACTTTGATTAATCTTGTAGCT	194
Qy	6406	GTATGCTCTTTTGAAGAAGCATTTGGGGAGTCTGATCCCTTTTGTATTTTAAATACAAATA	6465
Db	193	GTATGCTCTTTTGAAGAAGCATTTGGGGAGTCTGATCCCTTTTGTATTTTAAATACAAATA	134
Qy	6466	ATTGTACATATTTGCTATATTTTGTGTGAAGATGTGAAGATGTACTATGTTTATGCTTC	6525
Db	133	ATTGTACATATTTGCTATATTTTGTGTGAAGATGTGAAGATGTACTATGTTTATGCTTC	74
Qy	6526	TACATCCAGTTTGTACAAAGCTGGAATAATTAATATTAACAT	6568
Db	73	TACATCCAGTTTGTACAAAGCTGGAATAATTAATATTAACAT	31

RESULT 7					
AL048301/c					
LOCUS	810 bp	mRNA	linear	EST 01-MAR-2000	
DEFINITION	DKF7P586A1824.s1 586 (synonym: hutel) Homo sapiens cDNA clone				

DKFZp586a1824, mRNA sequence.  
 ACCESSION AL048301  
 VERSION AL048301.1 GI:4727441  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 810)  
 Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann  
 'S. (Duesterhoeft, et al.)  
 EST (Duesterhoeft, 1999)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Duesterhoeft A  
 MIPs  
 Am Klopferstr. 18a D-82152 Martinsried, Germany  
 This is the 3' sequence of the clone insert.  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
 sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.  
 No fl sequence available.  
 This clone (DKFZp586a1824) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
 FEATURES  
 source  
 1..810  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="DKFZp586a1824"  
 /clone\_1ib="586 (synonym: hute1)"  
 /tissue\_type="uterus"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Vector: pSPori; Site\_1: NotI; Site\_2: SalI/MluI"  
 BASE COUNT 285 a 138 c 135 g 245 t 7 others  
 ORIGIN  
 Query Match 10.5%; Score 690.4; DB 9; Length 810;  
 Best Local Similarity 95.9%; Pred. No. 3.1e-129;  
 Matches 748; Conservative 0; Mismatches 27; Indels 5; Gaps 4;

QY 6269 CAAGTGCACAAATGTTATCTATAGCATTAAGTAATTGTAGAACATAGAGACTGCTAATTC 6328  
 |||||||  
 Db 324 CAAGTGCACAAATGTTATCTATAGCATTAAGTAATTGTAGAACATAGAGACTGCTAATTC 265  
 QY 6329 AGTTCGCTGTGATGTCACAGTGCAGAAATGATACAAATTAAGTGTGATTCCTATACCTT 6388  
 |||||||  
 Db 264 AGTTCGCTGTGATGTCACAGTGCAGAAATGATACAAATTAAGTGTGATTCCTATACCTT 205  
 QY 6389 TGATTAAGTGTGATGTCACAGTGCAGAAATGATACAAATTAAGTGTGATTCCTATACCTT 6448  
 |||||||  
 Db 204 TGATTAAGTGTGATGTCACAGTGCAGAAATGATACAAATTAAGTGTGATTCCTATACCTT 145  
 QY 6449 GATATTAAGTGTGATGTCACAGTGCAGAAATGATACAAATTAAGTGTGATTCCTATACCTT 6508  
 |||||||  
 Db 144 GATATTAAGTGTGATGTCACAGTGCAGAAATGATACAAATTAAGTGTGATTCCTATACCTT 85  
 QY 6509 TACTATGTTATGCTTCTACATCCAGTTCGTACACAGCTGGAATAATTAATTAACAT 6568  
 |||||||  
 Db 84 TACTATGTTATGCTTCTACATCCAGTTCGTACACAGCTGGAATAATTAATTAACAT 25  
 RESULT 8  
 BM717124/C  
 LOCUS BM717124  
 DEFINITION UI-E-EJ0-abk-a-24-0-UI r2 UI-E-EJ0 Homo sapiens cDNA clone  
 UI-E-EJ0-abk-a-24-0-UI 5', mRNA sequence.  
 ACCESSION BM717124  
 VERSION BM717124  
 KEYWORDS EST.  
 SOURCE BM717124.1 GI:19030382  
 ORGANISM human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 697)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 9704477  
 COMMENT Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: msoares@blue.weeg.uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 CDNA library preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
 Seq primer: M13 Reverse.  
 FEATURES  
 source  
 1..697  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-EJ0-abk-a-24-0-UI"  
 /clone\_1ib="UI-E-EJ0"  
 /tissue\_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina foveal and Macular, RPE and Choroid"  
 /dev\_stage="fetal and adult"  
 /note="Organ: eye; Vector: pT73-Pac (pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-E-EJ0 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac

vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dt)<sub>18</sub> tail. The sequence tags for this library are: fetal eyes, AGAATCAAGA; lens, CGATTACGGA; eye anterior segment, AATGCCGAT; optic nerve, CCATTAGT; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).

BASE COUNT 167 a 153 c 150 g 227 t  
ORIGIN

Query Match 10.5%; Score 688; DB 14; Length 697;  
Best Local Similarity 100.0%; Pred. No. 9,8e-129; Indels 0; Gaps 0;

Matches 688; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2384 AATTTCAGAGAAACAAACAGCTGAAGAGATGATCATTTAAGCATTTTCACAGAT 2443  
|||||  
DB 688 AATTTCAGAGAAACAAACAGCTGAAGAGATGATCATTTAAGCATTTTCACAGAT 629  
QY 2444 AGCAGCTGACGTAGGAGATGCAAGAAATTTTAACTCAATGTTTGCATCATCAGTGCCT 2503  
|||||  
DB 628 AGCAGCTGACGTAGGAGATGCAAGAAATTTTAACTCAATGTTTGCATCATCAGTGCCT 569  
QY 2504 AAACCTGGACACAGTGGCAGACCTGGGAGAACTTCCCATTAATACGA 2563  
|||||  
DB 568 AAACCTGGACACAGTGGCAGACCTGGGAGAACTTCCCATTAATACGA 509  
QY 2564 AAAACATTTTCAAGATCTCCAGACCTGTTTGAATCTTCCAGAAATGSCAAATATCG 2623  
|||||  
DB 508 AAAACATTTTCAAGATCTCCAGACCTGTTTGAATCTTCCAGAAATGSCAAATATCG 449  
QY 2624 TAATGTTCTCAATAGTAAATATCTACAACTCCCATTAATGCTTATCCAGATTAACA 2663  
|||||  
DB 448 TAATGTTCTCAATAGTAAATATCTACAACTCCCATTAATGCTTATCCAGATTAACA 389  
QY 2684 AAAGATCTCACTCTCTCAGAGAAATGATCAAAAGTACGCGGTGTCATTT 2743  
|||||  
DB 388 AAAGATCTCACTCTCTCAGAGAAATGATCAAAAGTACGCGGTGTCATTT 329  
QY 2744 TGAGAGCTAGATGATGCAAAAGAAATTCGTCACGTTGGCGAATGGCTTAGTGAA 2803  
|||||  
DB 328 TGAGAGCTAGATGATGCAAAAGAAATTCGTCACGTTGGCGAATGGCTTAGTGAA 269  
QY 2804 CATGAGACCTGCGCTCATGTTTCAGAGACTCGGAGAAAGATGGCGGATTTGGGCTCT 2863  
|||||  
DB 268 CATGAGACCTGCGCTCATGTTTCAGAGACTCGGAGAAAGATGGCGGATTTGGGCTCT 209  
QY 2864 CAGCAGAGGTAGTACAATGACACAGTGTAGATGTTGCTCAGACAGGTGCTCATAAAA 2923  
|||||  
DB 208 CAGCAGAGGTAGTACAATGACACAGTGTAGATGTTGCTCAGACAGGTGCTCATAAAA 149  
QY 2924 GCGGATGAGTGTAGTCTTCTTCATGCAATGCCAAAAAGCTTATAGATGCGCAAAATGCG 2993  
|||||  
DB 148 GCGGATGAGTGTAGTCTTCTTCATGCAATGCCAAAAAGCTTATAGATGCGCAAAATGCG 89  
QY 2984 TCGAAAAGTGAAGAGTACCTTTTCCAAATTTGAGCTAGAAATGACGAGAGACTCTTCA 3043  
|||||  
DB 88 TCGAAAAGTGAAGAGTACCTTTTCCAAATTTGAGCTAGAAATGACGAGAGACTCTTCA 29  
QY 3044 GACATTTATCTCTGAGTGTGAGCCAGCA 3071  
|||||  
DB 28 GACATTTATCTCTGAGTGTGAGCCAGCA 1

RESULT 9  
BM423354 1064 bp mRNA linear EST 29-JAN-2002  
LOCUS BM423354  
DEFINITION AGENCOURT 6402148 NIH\_MGC\_41 Homo sapiens cDNA clone IMAGE:5516489  
5', mRNA sequence.  
ACCESSION BM423354  
VERSION BM423354.1 GI:18391579  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1064)  
AUTHORS NIH-MGC http://mhc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: DCTD/DTF  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
plate: LNCM2017 row: c column: 18  
High quality sequence stop: 799.  
Location/Qualifiers  
1..1064  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NIH\_MGC\_41"  
/clone\_1ib="5516489"  
/issue\_type="amelanotic melanoma, cell line"  
/lab\_host="PH10B (phage-resistant)"  
/note="Organ: skin; Vector: pORF7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using zap-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library"

BASE COUNT 278 a 256 c 272 g 244 t 14 others  
ORIGIN

Query Match 10.4%; Score 681.4; DB 13; Length 1064;  
Best Local Similarity 97.1%; Pred. No. 2e-127;  
Matches 725; Conservative 0; Mismatches 18; Indels 4; Gaps 3;

QY 3439 GGAAGGCTATCTTGGCTCCAGTGTCTGTGATATTTTTCAGATTCGTGTCACA 3498  
|||||  
DB 180 GGAAGGCTATCTTGGCTCCAGTGTCTGTGATATTTTTCAGATTCGTGTCACA 239  
QY 3499 GTGAATTTCTTCAAGATTCAGATTTGTTAGCAATTCGTTTGAAGTGCAGAGTCT 3558  
|||||  
DB 240 GTGAATTTCTTCAAGATTCAGATTTGTTAGCAATTCGTTTGAAGTGCAGAGTCT 299  
QY 3559 CACTGCAGATGAGAGGCGCGAGAGCATTCGTGACATCGTGGAAACAACTAGGGA 3618  
|||||  
DB 300 CACTGCAGATGAGAGGCGCGAGAGCATTCGTGACATCGTGGAAACAACTAGGGA 359  
QY 3619 TGGCCAGATGAGAGGCGCGAGCATTCGTGACATTCGTGAGTTCAGTTCAGTATG 3678  
|||||  
DB 360 TGGCCAGATGAGAGGCGCGAGCATTCGTGACATTCGTGAGTTCAGTTCAGTATG 419  
QY 3679 CACCAATGTCGAGAGGCGCGAGGCTTATATGCTAGACGTAGATTAATTTCTTCCAA 3738  
|||||  
DB 420 CACCAATGTCGAGAGGCGCGAGGCTTATATGCTAGACGTAGATTAATTTCTTCCAA 479  
QY 3739 CAGAGAACTTTCCAGAGATCAGGGGATCGCGGTGATGCTGCTGACAGTGGCC 3798  
|||||  
DB 480 CAGAGAACTTTCCAGAGATCAGGGGATCGCGGTGATGCTGCTGACAGTGGCC 539  
QY 3799 GTGGAGCTGAGAGTCAATGCTCAAGTGGCTCCCATGATTAATATACAGATCCAGACC 3858  
|||||  
DB 540 GTGGAGCTGAGAGTCAATGCTCAAGTGGCTCCCATGATTAATATACAGATCCAGACC 599  
QY 3859 AGAGAGCTGGAGAGCTTTCATTCGGGATTAATTCATTCATTCATTCATTCATTCATTC 3918  
|||||  
DB 600 AGAGAGCTGGAGAGCTTTCATTCGGGATTAATTCATTCATTCATTCATTCATTCATTC 659

QY 3919 CAGGTTTATGGGCATCAGCCATATGACCAATTAATGTTTCTGATCATAGCACAA 3978  
|||||  
Db 660 CAGGTTTATGGGCATCAGCCATATGACCAATTAATGTTTCTGATCATAGCACAA 719  
QY 3979 AGTATACAGGCAAAATCAAGTAGAGAGACCTTGAACAGCCCGAGCAAGCT 4038  
|||||  
Db 720 AGTATACAGGCAAAATCAAGTAGAGAGACCTTGAACAGCCCGAGCAAGCT 779  
QY 4039 GGGCGCTTCCACAGGTTCTGGGAGAGACTCAGAGGTGACAGGACACATTAAGC 4098  
|||||  
Db 780 GGGCGCTTCCACAGGTTCTGGGAGAGACTCAGAGGTGACAGGACACATTAAGC 839  
QY 4099 GGAGGGGTGGAA-AGGATGTTTCCATT--GAAGCGGAAAGACAGTACCTAACGCT-GTG 4154  
|||||  
Db 840 GGAGGGGTGGAAAGAGATGTTTCCATTGGAAGCCGACAGAGTACGCTTACGCTGGTG 899  
QY 4155 ACTACGGAAGAAACCAAGCCTGTGCCC 4181  
|||||  
Db 900 ACTACGGAAGAAACCAAGCCTGTGCCC 926

RESULT 10  
B0003665/c 703 bp mRNA linear EST 26-MAR-2002  
LOCUS  
DEFINITION UI-H-EII-aze-b-19-0-UI-s1 NCI\_CGAP\_EII Homo sapiens cDNA clone  
IMAGE:5847474 3', mRNA sequence.

ACCESSION B0003665  
VERSION B0003665  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 703)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bcrf.org

Journal COMMENT  
Tissue Procurement: Dr. Jose Mercuende  
cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be found  
through the I.M.A.G.E. Consortium/BLNI at: http://image.llnl.gov  
The following repetitive elements were found in this cDNA  
sequence: 1-36, >AT-rich#low\_complexity (matched complement)  
Seq primer: M13 FORWARD  
POLYA=yes.

FEATURES  
Source Location/Qualifiers

1..703  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5847474"  
/clone.lib="NCI\_CGAP\_EII"  
/tissue\_type="Chondrosarcoma"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia)  
with a modified polylinker; Site.1: EcoR I; Site.2: Not I;  
NCI\_CGAP\_EII is a normalized cDNA library containing the  
following tissue(s): Chondrosarcoma. The library was  
constructed according to Bonaldo, Lennon and Soares,  
Genome Research, 6:791-806, 1996. First strand cDNA  
synthesis was primed with an oligo-dT primer containing a  
Not I site. Double stranded cDNA was ligated to an EcoR I  
adaptor, digested with Not I, and cloned directionally  
into pT73-Pac vector. The oligonucleotide used to prime  
the synthesis of first-strand cDNA contains a library tag  
sequence that is located between the Not I site and the  
(dT)18 tail. The sequence tag for this library is

ACACTTGCAC.  
TAG\_LIB=UI-H-EII  
TAG\_TISSUE=chondrosarcoma  
TAG\_SEQ=ACACTTGCAC"  
BASE COUNT 258 a 118 c 107 g 220 t  
ORIGIN

Query Match 10.3%; Score 676; DB 14; Length 703;  
Best local similarity 99.9%; Pred. No. 2,6e-126;  
Matches 687; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 5881 GACTGAATGACCCCTCCACTATATTTTGTGTTTTCACAGACTCCGAAAAGTGA 5940  
|||||  
Db 703 GACTGAATGACCCCTCCACTATATTTTGTGTTTTCACAGACTCCGAAAAGTGA 644  
QY 5941 AGGCTGCCAATCTGAGTAGTACCAATGAGAGAACTGCTGGTCTTGATTTTTC 6000  
|||||  
Db 643 AGGCTGCCAATCTGAGTAGTACCAATGAGAGAACTGCTGGTCTTGATTTTTC 584  
QY 6001 ATTAATTCAGTGCATATATGATGATGATGATGATGATGATGATGATGATGAT 6060  
|||||  
Db 583 ATTAATTCAGTGCATATATGATGATGATGATGATGATGATGATGATGATGAT 524  
QY 6061 GTGGAATTCAGTGTGTTTTCACCTGATGATGATGATGATGATGATGATGATGAT 6120  
|||||  
Db 523 GTGGAATTCAGTGTGTTTTCACCTGATGATGATGATGATGATGATGATGATGAT 464  
QY 6121 CTTCGTATCATGCGATTTTCTGATGATGATGATGATGATGATGATGATGATGAT 6180  
|||||  
Db 463 CTTCGTATCATGCGATTTTCTGATGATGATGATGATGATGATGATGATGATGAT 405  
QY 6181 TTACATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6240  
|||||  
Db 404 TTACATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 345  
QY 6241 ATGAAAGACAGAAATCTAGCTTACACAAAGTGCACAAATGTTATTAAGATTAGT 6300  
|||||  
Db 344 ATGAAAGACAGAAATCTAGCTTACACAAAGTGCACAAATGTTATTAAGATTAGT 285  
QY 6301 AATTGTAGAACATAGAGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 6360  
|||||  
Db 284 AATTGTAGAACATAGAGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 225  
QY 6361 CAATTAACTGCTGATTTCCCATATCTTTGATGATGATGATGATGATGATGATGAT 6420  
|||||  
Db 224 CAATTAACTGCTGATTTCCCATATCTTTGATGATGATGATGATGATGATGATGAT 165  
QY 6421 AGACATTTGAGAGTCTGATCCCTTTTGTATTTTAAATCAATTAATGATGATGAT 6480  
|||||  
Db 164 AGACATTTGAGAGTCTGATCCCTTTTGTATTTTAAATCAATTAATGATGATGAT 105  
QY 6481 TATATTTTGTGAGAGTGTAGAAATGTACTATGTTTAACTTCAATCCAGTTTGA 6540  
|||||  
Db 104 TATATTTTGTGAGAGTGTAGAAATGTACTATGTTTAACTTCAATCCAGTTTGA 45  
QY 6541 CAGGCTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6568  
|||||  
Db 44 CAGGCTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 17

RESULT 11  
AL036264 702 bp mRNA linear EST 29-FEB-2000  
LOCUS  
DEFINITION DKF2564L14.2\_p1 564 (synonym: hbr2) Homo sapiens cDNA clone  
IMAGE:5847474 5', mRNA sequence.

ACCESSION AL036264  
VERSION AL036264  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 702)

AUTHORS Duesterhoeft, A., Lauber, J., Mewes, H.W., Gassenhuber, J. and Wiemann, S.  
 TITLE EST (Duesterhoeft, et al.)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Duesterhoeft A

MIPS  
 Am Kiofepersplatz 18a D-82152 Martinsried, Germany  
 This is the 5' sequence of the clone insert  
 clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;  
 sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.  
 No sl sequence available.  
 This clone (DKFZp564l142) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, Germany; Email: clone@rzpd.de.  
 Location/Qualifiers

## FEATURES

source  
 1..702  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="DKFZp564l142"  
 /clone\_id="564 (synonym: hibr2)"  
 /tissue\_type="brain"  
 /dev\_stage="fetal"  
 /lab\_host="X1-2blue"  
 /note="Vector: PAMPI; Site\_1: NotI; Site\_2: SalI"

## BASE COUNT

194 a 124 c 147 g 237 t

## ORIGIN

Query Match 9.9%; Score 648; DB 9; Length 702;  
 Best Local Similarity 98.4%; Pred. No. 1.2e-120;  
 Matches 665; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 5453 GAGCGAGAGCGGTGATGGCGGATGTCCTGCTCCATGAGATGGCAAAACCC 5512  
 1 GGGCGAGAGCGGTGATGGCGGATGTCCTGCTCCATGAGATGGCAAAACCC 60  
 Db 1 ATTTTAACTATATTTCTTTGATTTTGTATTTAGAGTGTGTTGTTTGT 5512  
 QY 5513 ATTTTAACTATATTTCTTTGATTTTGTATTTAGAGTGTGTTTGT 5512  
 1 ATTTTAACTATATTTCTTTGATTTTGTATTTAGAGTGTGTTTGT 120  
 Db 61 ATTTTAACTATATTTCTTTGATTTTGTATTTAGAGTGTGTTTGT 120  
 QY 5573 TTTTGTGTTTTTAAAGAAACATTTTAACTGATGATTCAGTGAAGCAGCT 5632  
 1 TTTTGTGTTTTTAAAGAAACATTTTAACTGATGATTCAGTGAAGCAGCT 180  
 Db 121 TTTTGTGTTTTTAAAGAAACATTTTAACTGATGATTCAGTGAAGCAGCT 180  
 QY 5633 GGGATGTGGAGCTAATGCCAGCTTTTATCTGCTCTTCAAGACGCTCCCTTAT 5692  
 1 GGGATGTGGAGCTAATGCCAGCTTTTATCTGCTCTTCAAGACGCTCCCTTAT 240  
 Db 181 GGGATGTGGAGCTAATGCCAGCTTTTATCTGCTCTTCAAGACGCTCCCTTAT 240  
 QY 5693 GAATTTGGCATTAGGGAATAACAAGCCTTTAAACGTGATAAAGATCAAAAACCTGGTTA 5752  
 241 GAATTTGGCATTAGGGAATAACAAGCCTTTAAACGTGATAAAGATCAAAAACCTGGTTA 300  
 Db 5753 GACATGGCAGCTTTGCAAGGAGTTAGTACCAACAAAGACTCAAGTGGCTTAT 5812  
 1 GACATGGCAGCTTTGCAAGGAGTTAGTACCAACAAAGACTCAAGTGGCTTAT 360  
 Db 301 GACATGGCAGCTTTGCAAGGAGTTAGTACCAACAAAGACTCAAGTGGCTTAT 360  
 QY 5813 GGAGCGTCATATAGAGAGGCTTAAGTGTAGCAACATCTGCTCACAGCTGATTAAC 5872  
 361 GGAGCGTCATATAGAGAGGCTTAAGTGTAGCAACATCTGCTCACAGCTGATTAAC 420  
 Db 5873 CCTATATGACTGAATGAGACCCCTCCACTATTTTGTGTTTGGACAGACTCCG 5932  
 421 CCTATATGACTGAATGAGACCCCTCCACTATTTTGTGTTTGGACAGACTCCG 480  
 QY 5933 AAAAGTGAAGCTGCCATCTGAGTACTCAATGTGAGAACTGCTGCTGGATT 5992  
 481 AAAAGTGAAGCTGCCATCTGAGTACTCAATGTGAGAACTGCTGCTGGATT 540  
 Db 5993 TTTTTCATTAAATTCAGTGTATTTGATCAGTGAAGATGAAGATTAAGCTTCAA 6052  
 541 TTTTTCATTAAATTCAGTGTATTTGATCAGTGAAGATGAAGATTAAGCTTCAA 600

QY 6053 TTTTAAAGTGAATTCAGTGTGTTTTCAGTATCAAAACATGTCAGTCTTATTTA 6112  
 1 TTTTAAAGTGAATTCAGTGTGTTTTCAGTATCAAAACATGTCAGTCTTATTTA 659  
 Db 601 TTTTAAAGTGAATTCAGTGTGTTTTCAGTATCAAAACATGTCAGTCTTATTTA 659  
 QY 6113 ATTAATTCCTCTGCTGA 6128  
 1 ATTAATTCCTCTGCTGA 675  
 Db 660 ATTAATTCCTCTGCTGA 675

## RESULT 12

BM638008 646 bp mRNA linear EST 06-MAR-2002

LOCUS K-EST0114179 S9SNU601 Homo sapiens cDNA clone S9SNU601-74-All 5',  
 DEFINITION mRNA sequence.

ACCESSION BM638008.1 GI:19194417

VERSION BM638008.1  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 646)  
 Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,  
 Oh, K.U., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and  
 Kim, Y.S.

AUTHORS 21C Frontier Korean EST Project 2001  
 JOURNAL Unpublished (2002)  
 CONTACT: Kim YS

COMMENT Genome Research Center  
 Korea Research Institute of Bioscience & Biotechnology  
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
 Tel: +82-42-860-4470  
 Fax: +82-42-860-4409  
 Email: yongsung@email.kribb.re.kr  
 Plate: 74 row: A column: 11  
 High quality sequence stop: 646.

FEATURES  
 source  
 1..646  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="S9SNU601-74-All"  
 /clone\_lib="S9SNU601"  
 /sex="M"  
 /tissue\_type="ascites"  
 /cell\_type="Epithelial"  
 /cell\_line="SNU-601"  
 /lab\_host="Top10F"  
 /note="Organ: Stomach; Vector: pME18-FL3; Site\_1: XhoI;  
 Site\_2: XhoI; The poly (A)+ RNA was dephosphorylated with  
 bacterial alkaline phosphatase (BAP) and then deapped  
 with tabacco acid pyrophosphatase (TAP). The deapped  
 intact mRNA was ligated with DNA-RNA linker including SfiI  
 site by treatment of 74 RNA ligase and the first strand  
 cDNA was synthesized with Superscript II using SfiI  
 oligo-dT primer. After first strand synthesis, RNA was  
 degraded by NaOH treatment and cDNA was amplified by PCR  
 reaction. The PCR products were digested with SfiI and  
 cloned into DraIII- digested pME18S-FL3 vector. The  
 obtained cDNA vectors were used for transformation of  
 competent cells E. coli Top10F by electroporation method.  
 The cDNA libraries constructed by this method are  
 full-length enriched cDNA library."

BASE COUNT 220 a 104 c 144 g 178 t

ORIGIN  
 Query Match 9.8%; Score 646; DB 14; Length 646;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-120;  
 Matches 646; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 903 ATGCATTGCTGGGAAGAGCATTCAGTACTAGATCCACATTCATAGAAAGACTTCTGTG 962  
 1 ATGCATTGCTGGGAAGAGCATTCAGTACTAGATCCACATTCATAGAAAGACTTCTGTG 60  
 Db 1 ATGCATTGCTGGGAAGAGCATTCAGTACTAGATCCACATTCATAGAAAGACTTCTGTG 60



```

QY 963 ACCTATGAGACTTTCTTCTTCTACGCCCATGGAATGGGCAAAAAGTTATTGAGTGT 1022
    |||||||
Db 61 ACCTATGAGACTTTCTTCTTCTTCTACGCCCATGGAATGGGCAAAAAGTTATTGAGTGT 120
QY 1023 AATGACCCGAGCCTCAGGGATTAAGTTACACGGGTAGTATTATTGGGTAATAATCAC 1082
    |||||||
Db 121 AATGACCCGAGCCTCAGGGATTAAGTTACACGGGTAGTATTATTGGGTAATAATCAC 180
QY 1083 TTCAATGACTTTGAAGAGATCTCTGCATGACATCGAATTTTAAAGAAATTAACAAT 1142
    |||||||
Db 181 TTCAATGACTTTGAAGAGATCTCTGCATGACATCGAATTTTAAAGAAATTAACAAT 240
QY 1143 CTGGAAGAAGAGAAATGGGTGAGACACTAAGCCTGTTGAATATCGCGTGTCTGCTAAA 1202
    |||||||
Db 241 CTGGAAGAAGAGAAATGGGTGAGACACTAAGCCTGTTGAATATCGCGTGTCTGCTAAA 300
QY 1203 GCAGAAAAGAGATTGATGACGTTAACAACCAATCCCGAGACCTCTTGGCTTTATC 1262
    |||||||
Db 301 GCAGAAAAGAGATTGATGACGTTAACAACCAATCCCGAGACCTCTTGGCTTTATC 360
QY 1263 TTACTTGGAGGCTCTGTGAAGGGATTTGAATCTTTGTTGACATGATTCAGGTAGC 1322
    |||||||
Db 361 TTACTTGGAGGCTCTGTGAAGGGATTTGAATCTTTGTTGACATGATTCAGGTAGC 420
QY 1323 AAGCAACTGAGCAGGCTTGAAACGGGGGATCAGATATTAGAAGTAATGGCCAAAC 1382
    |||||||
Db 421 AAGCAACTGAGCAGGCTTGAAACGGGGGATCAGATATTAGAAGTAATGGCCAAAC 480
QY 1383 TTGGAAGAACCTTCACTGTCTAAAGCTATGGAATCTTGAAGTAACACACATTTATCT 1442
    |||||||
Db 481 TTGGAAGAACCTTCACTGTCTAAAGCTATGGAATCTTGAAGTAACACACATTTATCT 540
QY 1443 ATACTGTGAAACCAATTTATTTGTTTAAAGAACTTCAACAGATTTGTCAGAAAG 1502
    |||||||
Db 541 ATACTGTGAAACCAATTTATTTGTTTAAAGAACTTCAACAGATTTGTCAGAAAG 600
QY 1503 AAAAGAATGTGCCCCCACCCTTCTCTAAATTTGTGACATTAATA 1548
    |||||||
Db 601 AAAAGAATGTGCCCCCACCCTTCTCTAAATTTGTGACATTAATA 646

RESULT 13
BM700143 645 bp mRNA linear EST 28-FEB-2002
LOCUS UI-E-DWI-ahc-e-09-0-UI.r1 UI-E-DWI Homo sapiens cDNA clone
DEFINITION UI-E-DWI-ahc-e-09-0-UI 5', mRNA sequence.
ACCESSION BM700143
VERSION BM700143.1 GI:19013401
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 645)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLES Normalization and subtraction: two approaches to facilitate gene
discovery.
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.

```

```

FEATURES
    source          Location/Qualifiers
                    1..645
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="UI-E-DWI-ahc-e-09-0-UI"
                    /clone_lib="UI-E-DWI"
                    /tissue_type="lens"
                    /dev_stage="adult"
                    /lab_host="PH10B (life technologies) (71 phage resistant)"
                    /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
                    modified polylinker; Site_1: EcoR I; Site_2: Not I;
                    UI-E-DWI is a normalized cDNA library containing the
                    following tissue(s): lens. The library was constructed
                    according to Bonaldo, Lennon and Soares, Genome Research,
                    5:791-806, 1996. First strand cDNA synthesis was primed
                    with an oligo-dT primer containing a Not I site. Double
                    stranded cDNA was ligated to an EcoR I adaptor, digested
                    with Not I, and cloned directionally into pT73-Pac
                    vector. The oligonucleotide used to prime the synthesis of
                    first-strand cDNA contains a library tag sequence that is
                    located between the Not I site and the (dT)18 tail. The
                    sequence tag for this library is CGATTAGCA. This library
                    was created for the program, Gene Discovery in the Visual
                    System, supported by National Eye Institute (NEI)."
```

BASE COUNT 172 a 190 c 171 g 111 t 1 others

ORIGIN

```

Query Match          9.88; Score 642.4; DB 14; Length 645;
Best Local Similarity 99.7%; Pred. No. 1.7e-119;
Matches 643; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4007 GAGCCTTGAAACAGCCAGCCAGCCGAGCAGCTGGGCTCTTCCACAGTTACTGGGAGA 4066
    |||||||
Db 1 GAGCCTTGAAACAGCCAGCCAGCTGGGCTCTTCCACAGTTACTGGGAGA 60
QY 4067 AGACTGAAAGGTGACACAGCAGCAATTAAGCGAGGGGTGGAAGGATTTCCATTGA 4126
    |||||||
Db 61 AGACTGAAAGGTGACACAGCAGCAATTAAGCGAGGGGTGGAAGGATTTCCATTGA 120
QY 4127 AGCCGAAGCAGTAGTACCTTAACGTCGTGACTACGGAAGAAACCAAGCTGTCCCATGCC 4186
    |||||||
Db 121 AGCCGAAGCAGTAGTACCTTAACGTCGTGACTACGGAAGAAACCAAGCTGTCCCATGCC 180
QY 4187 TGCCCATATGCTGTGACATCAAGTACTACAAAGGGGCTCATTCAGAAAGGAGGCGAG 4246
    |||||||
Db 181 TGCCCATATGCTGTGACATCAAGTACTACAAAGGGGCTCATTCAGAAAGGAGGCGAG 240
QY 4247 GATCGAGAGCCCCCGCCCAACCCCTCCGGCTACATTGGAATTTCCATTACTGACTTCC 4306
    |||||||
Db 241 GATCGAGAGCCCCCGCCCAACCCCTCCGGCTACATTGGAATTTCCATTACTGACTTCC 300
QY 4307 AGAAGGACATCCCATATCCATCCAGCAGGAACCGCGGACTACAGCTGGCCCTTCAGACATC 4366
    |||||||
Db 301 AGAAGGACATCCCATATCCATCCAGCAGGAACCGCGGACTACAGCTGGCCCTTCAGACATC 360
QY 4367 GCGGATGTGACAGATCTCCGACACAGCTGGGCTTCATCCGTACAGAGCCACATGG 4426
    |||||||
Db 361 GCGGATGTGACAGATCTCCGACACAGCTGGGCTTCATCCGTACAGAGCCACATGG 420
QY 4427 GCATCCGACAGCAGCAGGCTGTGAACAAACCTCAGTGGCATTAACGGAAGAGTCTGA 4486
    |||||||
Db 421 GCATCCGACAGCAGCAGGCTGTGAACAAACCTCAGTGGCATTAACGGAAGAGTCTGA 480
QY 4487 CCGGCGCTCGCCCTTATATAGTCCCAAGGGTTTCCACGAGAGAGATGAATGACA 4546
    |||||||
Db 481 CCGGCGCTCGCCCTTATATAGTCCCAAGGGTTTCCACGAGAGAGATGAATGACA 540
QY 4547 AGTTTGTGTTTGAGGACAGACTTTCTGGAAGAGAGCGGACCGCAAGGAAGAGA 4606
    |||||||
Db 541 AGTTTGTGTTTGAGGACAGACTTTCTGGAAGAGAGCGGACCGCAAGGAAGAGA 600
QY 4607 GCACAAAGACGTCCTGACCATTTGAGGCTTGGAACTCACTATTC 4651
    |||||||

```

Db	601	GCACANGAAGACGCTCTGACATTTGGAGCCTTGGAACTGCATTC	645
RESULT_14	AM467230	643 bp	mRNA
LOCUS	he07906.x1	NCI CGAP CML1	Homo sapiens CDNA clone IMAGE:2918362 3'
DEFINITION	similar to TR:043386	043386 KIA001LB	;; mRNA sequence.
ACCESSION	AM467230		
VERSION	AM467230.1	GI:7037336	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 643)		
AUTHORS	NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap.	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
COMMENT	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cegp@strfemail.nih.gov		
	Tissue Procurement: Elisabeth Paietta, Jonathan D. Licht, M.D.,		
	Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life		
	Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The		
	I.M.A.G.E. Consortium DNA Sequencing by: Washington University		
	Genome Sequencing Center		
	Clone distribution: NCI-CGAP clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	www.bio.llnl.gov/bbrp/image/image.html		
	Seq primer: -400P from G1bco		
	High quality sequence stop: 391.		
FEATURES	Location/Qualifiers		
source	1..643		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:2918362"		
	/clone_1lb="NCI-CGAP CML1"		
	/tissue_type="myeloid cells; 18 pooled CML cases, BCR/ABL		
	rearrangement positive, includes both chronic phase and		
	myeloid blast crisis"		
	/lab_host="DH10B"		
	/note="Organ: Whole blood; Vector: PCMV-SPORT6; Site_1:		
	Sal1; Site_2: NotI; cloned unidirectionally. Primer:		
	Oligo dt. Library constructed by Life Technologies."		
BASE COUNT	152 a 172 c 129 g 190 t		
ORIGIN			
Query Match	9.5%; Score 627; DB 10; Length 643;		
Best Local Similarity	98.4%; Pred. No. 2..2e-116;		
Matches	633; Conservative 0; Mismatches 10; Indels 0; Gaps 0		
QY	147 CATCTTACTGACAGTCCTCCACCCACAGGTGACGCTTCTTCTGACCATTCAGATGT	206	
DB	643 CATCTTACTGACAGTCCTCCACCCACAGGTGACGCTTCTTCTGACCATTCAGATGT	584	
QY	207 AGTATCACTAGTGAATTCGGGAGCAGCAGTCTTCTGATATCTTACACGGCCACAGAAAGC	266	
DB	583 AGTATCACTAGTGAATTCGGGAGCAGCAGTCTTCTGATATCTTACACGGCCACAGAAAGC	524	
QY	267 GAGGCTGTGATATGACCTGACGTGCGTGGCCAGAAACAGCAGTTCGGAAGACGAC	326	
DB	523 GAGGCTGTGATATGACCTGACGTGCGTGGCCAGAAACAGCAGTTCGGAAGACGAC	464	
QY	327 GACGATGAAGACATTCGAGAGCATCAATCCTCTGATGAGCAGGACATTTGTGAGA	386	
DB	463 GACGATGAAGACATTCGAGAGCATCAATCCTCTGATGAGCAGGACATTTGTGAGA	404	
QY	387 GACTGCTAGAGAGACCAATTGACCGGACAGATGATGACATTTGAACAACCTGTGAA	446	
DB	403 GACTGCTAGAGAGACCAATTGACCGGACAGATGATGACATTTGAACAACCTGTGAA	344	
QY	447 TTTATGACACAGTGGCTCTTTTGCAATATGACATGTGAGTGGCGAGAACTCTGT	506	

Db	343	TTTATGACCCAGTTGCCCTGGCTTTTGCCCAATTGACANTGTCACTGAGCGSAGAACTCTGT	284
OY	507	GCTGTGATGAGTCGTTCCGCAGTGTGGAAAGAGCAGGACCATACTAGTGTAAAATGATGTTGA	566
Dd	283	GCTGTGATGAGTCGTTCCGCAGTGTGGAAAGAGCAGGACCATACTAGTGTAAAATGATGTTGA	224
OY	567	GAGTCGACCTCCTGTCAGTATTTCTCAATGAGATCTGGAAAGCATTTTCCAGATGGA	626
Dd	223	GAGTCGACCTCCTGTCAGTATTTCTCAATGAGATCTGGAAAGCATTTTCCAGATGGA	164
OY	627	AAGAGCAAATTACTGTGCATGGGAAATAGTTTGGTGTCTCTCTCCATCAGACAAGAA	866
Dd	163	AAGAGCAAATTACTGTGCATGGGAAATAGTTTGGTGTCTCTCTCCATCAGACAAGAA	104
OY	687	TACATCAAAAGAGTAGTATGAGAACAAAGTGCATACGCCAGTTTGTGTGATPAGCCAG	746
Dd	103	TACATCAAAAGAGTAGTATGAGAACAAAGTGCATACGCCAGTTTGTGTGATPAGCCAG	44
OY	747	CAGATTACTGCCGATTCTCATCACTAAGTAGAAAAAGACATGC	789
Dd	43	CAGATTACTGCCGATTCTCATCACTAAGTAGAAAAAGACATGC	1
RESULT 15			
AL596512		625 bp mRNA linear EST 14-AUG-2001	
LOCUS		DKFP451B1910.r1_451 (synonym: hlcc1) spinal cord Homo sapiens cDNA	
DEFINITION		clone DKFP451B1910 5', mRNA sequence.	
ACCESSION		AL596512	
VERSION		AL596512.1 GI:15154208	
KEYWORDS		EST.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
TITLE		1 (bases 1 to 625)	
JOURNAL		Ottewaelder,B., Obermaier,B., Mewes,W., Mewes,H.W., Well,B. and Wiemann,S.	
COMMENT		EST (Ottewaelder,B., Obermaier,B., Mewes,H.W., Well,B. and Wiemann,S.)	
		Unpublished (2001)	
		Contact: ottewaelder B	
		MPS	
		Am Klopferpfitz 18a D-82152 Martinsried, Germany	
		This is the 5' sequence of the clone insert	
		this is the 5' sequence of the clone insert	
		Clone from S. Wiemann, Molecular Genome Analysis, German Cancer	
		Research Center (DKFZ), Email:S.wiemann@dkfz-heidelberg.de;	
		Sequenced by Medigenomix (Marti.S.wiemann@germany) within the cDNA	
		sequencing consortium of the German Genome Project. No sl sequence	
		available.	
		This clone (DKFP451B1910) is available at the RZPD in Berlin.	
		Please contact the RZPD: Ressourcenzentrum, Heubenerweg 6, 14059	
		Berlin-Charlottenburg, GERMANY, Email: clone@rzpd.de.	
FEATURES		Location/Qualifiers	
source		1..625	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="DKFP451B1910"	
		/clone_lib="451 (synonym: hlcc1) spinal cord"	
		/tissue_type="human spinal cord"	
		/dev_stage="adult"	
		/lab_host="DH10B"	
		/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"	
BASE COUNT		212 a 156 g 159 t 1 others	
ORIGIN			
Query Match		9.5%; Score 624; DB 9; Length 625;	
Best Local Similarity		99.8%; Fred.No. 8.7e-116; Indels 0; Gaps 0	
Matches 624; Conservative		0; Mismatches 1;	
OY	676	TGACCAAGAAATACATGAAGAGATGATGAGAACAAAGGTGATGACTGCAGTTGTCT	735

```
Db      1  TGGACAAAGATACATGAAGAAGTAGATGAGAACAAAGGTGATGACTGCCAGTTTGTCT 60
QY      736  GCATAGCCCGCAGCAGATTACTGCCGATTTCTCAATCAGTAGAAAAAGCATGCAAAAAG 795
      |||
Db      61  GCATAGCCCGCAGCAGATTACTGCCGATTTCTCAATCAGTAGAAAAAGCATGCAAAAAG 120
QY      796  TTGAAGAGAGAGAGAGATTGTATGTGTAAAGAACCCGAGAACTTGATCGAACTGAA 855
      |||
Db      121  TTGAAGAGAGAGAGATTGTATGTGTAAAGAACCCGAGAACTTGATCGAACTGAA 180
QY      856  CAAGAAAGGAGACATCTGATCATCAAGGTACTCCAGAAAAGTTAACATGCAATTTGTGG 915
      |||
Db      181  CAAGAAAGGAGACATCTGATCATCAAGGTACTCCAGAAAAGTTAACATGCAATTTGTGG 240
QY      916  AAGAGCATTCAGTAGTAGATCCACATCATATAGAAAGCTTTCTGTGACCTATAGACTT 975
      |||
Db      241  AAGAGCATTCAGTAGTAGATCCACATCATATAGAAAGCTTTCTGTGACCTATAGACTT 300
QY      976  TTCTTTCTAGCCCAATGAAAGTGGCCAAAAGTTATTGGAGTGTAAATGACCCGAGCC 1035
      |||
Db      301  TTCTTTCTAGCCCAATGAAAGTGGCCAAAAGTTATTGGAGTGTAAATGACCCGAGCC 360
QY      1036  TCAGGATTAAGGTTACACGGGTAGTATTAATGTGGTAAATAATCACTTCATGACTTGG 1095
      |||
Db      361  TCAGGATTAAGGTTACACGGGTAGTATTAATGTGGTAAATAATCACTTCATGACTTGG 420
QY      1096  AAGGAGATCTCGAATGATCGATTTTGTAGAAATTTGAAAACATCTGGAAGAGAGA 1155
      |||
Db      421  AAGGAGATCTCGAATGATCGATTTTGTAGAAATTTGAAAACATCTGGAAGAGAGA 480
QY      1156  AATGGGTGGACACCTAAGGCTGTGAATATGCGCTGTCTGCTAAAGCAAAAAGAGAT 1215
      |||
Db      481  AATGGGTGGACACCTAAGGCTGTGAATATGCGCTGTCTGCTAAAGCAAAAAGAGAT 540
QY      1216  TGATGACGTTAACAAACCATCCCGAAGCTCTTTGCCCTTTATCTTACTTGGAGGCT 1275
      |||
Db      541  TGATGACGTTAACAAACCATCCCGAAGCTCTTTGCCCTTTATCTTACTTGGAGGCT 600
QY      1276  CTGAGAAGGAGATTGGATCTTTGT 1300
      |||
Db      601  CTGAGAAGGAGATTGGATCTTTGT 625
```

Search completed: February 18, 2003, 17:34:17  
Job time : 7868 secs

**BLANK PAGE**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 15:52:41 : Search time 330 Seconds  
(without alignments)  
10137.084 Million cell updates/sec

Title: US-09-911-826a-1

Perfect score: 6568

Sequence: 1 cttgcgcacgtgagagattg.....aaataataataataacat 6568

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCT05\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6568	100.0	6568	10	US-09-880-107-1535
2	6568	100.0	6568	10	US-09-911-826a-1
3	1587	24.2	6383	10	US-09-728-952-68
4	399	6.1	399	9	US-10-046-935-1879
5	399	6.1	399	9	US-09-878-178-1879
6	394.8	6.0	799	10	US-09-911-826a-3
7	388	5.9	400	9	US-10-046-935-1196
8	388	5.9	400	9	US-09-878-178-1196
9	227.8	3.5	263	10	US-09-733-607-15
10	201.6	3.1	360	10	US-09-864-761-27564
11	166.6	2.5	465	10	US-09-864-761-10928
12	104.6	1.6	5900	10	US-09-954-456-1970
13	104.6	1.6	5900	9	US-09-822-846-129
14	95.8	1.5	2536	9	US-09-864-761-22078
15	87.8	1.3	182	10	US-09-864-761-31466
16	72.8	1.1	181	10	US-09-864-761-14576
17	72.8	1.1	463	10	US-09-864-761-14576
18	60.8	0.9	4770	10	US-09-801-368-63
19	56.2	0.9	446	10	US-09-864-761-11216

C 20	52.4	0.8	383	10	US-09-864-761-28016	Sequence 28016, A
C 21	50	0.8	6200	9	US-09-993-241-1	Sequence 1, Appli
C 22	50	0.8	6200	10	US-09-993-008-1	Sequence 1, Appli
C 23	49.4	0.8	311	10	US-09-983-965-1973	Sequence 1973, Ap
C 24	49.2	0.7	2797	9	US-10-078-770-133	Sequence 133, App
C 25	49	0.7	1178	9	US-09-764-868-410	Sequence 410, App
C 26	49	0.7	1926	10	US-09-764-853-238	Sequence 238, App
C 27	49	0.7	2086	10	US-09-728-952-74	Sequence 74, Appl
C 28	49	0.7	6336	10	US-09-964-824a-114	Sequence 114, App
C 29	49	0.7	6336	10	US-09-880-107-1537	Sequence 1537, Ap
C 30	48.2	0.7	317	10	US-09-960-352-14196	Sequence 14196, A
C 31	48.2	0.7	1586	10	US-09-745-763-86	Sequence 86, Appl
C 32	48	0.7	3275	9	US-09-854-133-151	Sequence 151, App
C 33	48	0.7	3375	10	US-09-738-693-151	Sequence 151, App
C 34	47.6	0.7	337	10	US-09-764-846-100	Sequence 100, App
C 35	47.2	0.7	597	10	US-09-925-300-87	Sequence 87, Appl
C 36	47	0.7	545	10	US-09-764-873-12	Sequence 12, Appl
C 37	47	0.7	6200	9	US-10-206-443-1	Sequence 1, Appli
C 38	46.8	0.7	401	10	US-09-960-352-14060	Sequence 14060, A
C 39	46.8	0.7	416	10	US-09-960-352-4584	Sequence 4584, Ap
C 40	46.6	0.7	332	10	US-09-960-352-13112	Sequence 13112, A
C 41	46.6	0.7	1419	10	US-09-940-836a-3	Sequence 3, Appli
C 42	46.6	0.7	2807	10	US-09-940-836a-1	Sequence 1, Appli
C 43	46.4	0.7	203	10	US-09-960-352-3742	Sequence 3742, Ap
C 44	46.4	0.7	392	10	US-09-960-352-2634	Sequence 2634, Ap
C 45	46.2	0.7	219	10	US-09-864-761-27841	Sequence 27841, A

#### ALIGNMENTS

RESULT 1									
US-09-880-107-1535									
Sequence 1535, Application US/09880107									
Patent No. US20020142981A1									
GENERAL INFORMATION:									
APPLICANT: Horne, Darci T.									
APPLICANT: Vockley, Joseph G.									
APPLICANT: Scherif, Uwe.									
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer									
FILE REFERENCE: 44921-5028-WO									
CURRENT APPLICATION NUMBER: US/09/880,107									
CURRENT FILING DATE: 2001-06-14									
PRIOR APPLICATION NUMBER: US 60/211,379									
PRIOR FILING DATE: 2000-06-14									
PRIOR APPLICATION NUMBER: US 60/237,054									
PRIOR FILING DATE: 2000-10-02									
NUMBER OF SEQ ID NOS: 3950									
SOFTWARE: PatentIn Ver. 2.1									
SEQ ID NO 1535									
LENGTH: 6568									
TYPE: DNA									
ORGANISM: Homo sapiens									
FEATURE:									
OTHER INFORMATION: Genbank Accession No. US20020142981A1 AB002311									
US-09-880-107-1535									
Query Match									
Best Local Similarity 100.0%; Score 6568; DB 10; Length 6568;									
Matches 6568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	CTTGCCATCGTGTGAGACATTTGGTACATGATGTGTAATTCAGTTCAGCATATGTTCTTCA	60						
DB	1	CTTGCCATCGTGTGAGACATTTGGTACATGATGTGTAATTCAGTTCAGCATATGTTCTTCA	60						
QY	61	TTATGAACACACATGACATCCCGATACATGAGATTATGGGACAGAGAAACACT	120						
DB	61	TTATGAACACACATGACATCCCGATACATGAGATTATGGGACAGAGAAACACT	120						
QY	121	CACTTCCTGAGATTTCACAAACCTGACACTCTTCACACCCAGGTGACCC	180						
DB	121	CACTTCCTGAGATTTCACAAACCTGACACTCTTCACACCCAGGTGACCC	180						

181 ACCTCTCTAGCCATTGAGATGATAGTATCTAGTATTCCTGGAGCAGACGCTCTT 240  
181 ACGTTCTCTTAGCCATTGAGATGATAGTATCTAGTATTCCTGGAGCAGACGCTCTT 240  
241 CTGATATCTACAGCCACAGAAAGCAGGCTGGTATGAGACCTGAGGGCTTCCAG 300  
241 CTGATATCTACAGCCACAGAAAGCAGGCTGGTATGAGACCTGAGGGCTTCCAG 300  
301 AAACAGCAGTGGATTCGAAAGCAGCAGATGAAAGACATTTGAGAGCATTCAGATC 360  
301 AAACAGCAGTGGATTCGAAAGCAGCAGATGAAAGACATTTGAGAGCATTCAGATC 360  
361 CTCTGATGAGCAGGACATTTGTGAGAGACCTGCTGAGAGAGGACCCCAATTGACCAGCAG 420  
361 CTCTGATGAGCAGGACATTTGTGAGAGACCTGCTGAGAGAGGACCCCAATTGACCAGCAG 420  
421 ATGATGACATTTGAACACTCTTGGATTTATGCACAGTTGCCCTGCTTTGCCAATTTGA 480  
421 ATGATGACATTTGAACACTCTTGGATTTATGCACAGTTGCCCTGCTTTGCCAATTTGA 480  
481 CAATGTCAGTGAAGCAGAACTCTGCTGTGATGCTGTGCGACGTGGTGAAGAGCAG 540  
481 CAATGTCAGTGAAGCAGAACTCTGCTGTGATGCTGTGCGACGTGGTGAAGAGCAG 540  
541 GGCACCTAGTGTAAATGATGTTGTAAGAGCTGACTGCTGCTGATGATTTCTCAATGAT 600  
541 GGCACCTAGTGTAAATGATGTTGTAAGAGCTGACTGCTGCTGATGATTTCTCAATGAT 600  
601 CTGTGAGACGATTTATCCAGATGGAAGACAAATACTGTGATGGGAAATGTTTG 660  
601 CTGTGAGACGATTTATCCAGATGGAAGACAAATACTGTGATGGGAAATGTTTG 660  
661 GTGTCTCTCTACCATGAGCAAAAGATATCATGAAAGAGTGTGATGAAACCAAGTGATG 720  
661 GTGTCTCTCTACCATGAGCAAAAGATATCATGAAAGAGTGTGATGAAACCAAGTGATG 720  
721 ACTGCCAGTTTGTCTGCATAGCCACCAAGATTTACTGCTGATTTCTCATCAAGTAGAA 780  
721 ACTGCCAGTTTGTCTGCATAGCCACCAAGATTTACTGCTGATTTCTCATCAAGTAGAA 780  
781 AGAATGATGCAAAAGTTGAAAGAGAGAGATGTTATGTTGTAAGAAACCCGAGAAC 840  
781 AGAATGATGCAAAAGTTGAAAGAGAGAGATGTTATGTTGTAAGAAACCCGAGAAC 840  
841 TTGATGCAAACTGAAAGAGAGAGACATTTGTCATCAAGGTTACTCAGAAAGGTTAA 900  
841 TTGATGCAAACTGAAAGAGAGAGACATTTGTCATCAAGGTTACTCAGAAAGGTTAA 900  
901 CAATGATGTTGGTGAAGAGACATTCAGTATGATCCAACTTCATAGAGACTTTCTGT 960  
901 CAATGATGTTGGTGAAGAGACATTCAGTATGATCCAACTTCATAGAGACTTTCTGT 960  
961 TGACCTATGAGACTTTTCTTCTAGCCCAATGAGTGGGCAAAAAGTTATTTGAGAGGT 1020  
961 TGACCTATGAGACTTTTCTTCTAGCCCAATGAGTGGGCAAAAAGTTATTTGAGAGGT 1020  
1021 TTAATGACCCGAGCTCAGGGATAGGTTACAGGGTATGTTATTTGTTGGTAAATATC 1080  
1021 TTAATGACCCGAGCTCAGGGATAGGTTACAGGGTATGTTATTTGTTGGTAAATATC 1080  
1081 ACTTCAATACCTTTGAGAGAGATCTCTGCAATGACATCGATTTTGAAGAAATTTGAAACA 1140  
1081 ACTTCAATACCTTTGAGAGAGATCTCTGCAATGACATCGATTTTGAAGAAATTTGAAACA 1140  
1141 ATCTGGAAGAGAGAAATGGTGAACACTTAAGCTGTGTAATATGCGGCTGCTGCTA 1200  
1141 ATCTGGAAGAGAGAAATGGTGAACACTTAAGCTGTGTAATATGCGGCTGCTGCTA 1200  
1201 AAGCAAAAGAGATGATGATTAACAAACCATCCCGAGAGCTCTTGGCTTTTA 1260  
1201 AAGCAAAAGAGATGATGATTAACAAACCATCCCGAGAGCTCTTGGCTTTTA 1260

1261 TCTTACTTGGAGGCTCTGAGACAGGATTTTGAATCTTTGTGACATGATTCAGSTA 1320  
1261 TCTTACTTGGAGGCTCTGAGACAGGATTTTGAATCTTTGTGACATGATTCAGSTA 1320  
1321 GCAAGCACTGAAAGCAGGCTTGAACGGGGGATGATATTTAGAAAGTAAATGCGCAAA 1380  
1321 GCAAGCACTGAAAGCAGGCTTGAACGGGGGATGATATTTAGAAAGTAAATGCGCAAA 1380  
1381 ACTTGAAGCACTGAGCTGTCAAAAGCTATGCAATTTCTTGAATTAACACATTTAT 1440  
1381 ACTTGAAGCACTGAGCTGTCAAAAGCTATGCAATTTCTTGAATTAACACATTTAT 1440  
1441 CTATCAGTGAAGCAATTTATTTGATTTAAGAACTTTACAGAGATTCGAGAG 1500  
1441 CTATCAGTGAAGCAATTTATTTGATTTAAGAACTTTACAGAGATTCGAGAG 1500  
1501 AGAAAGAAATGTTGCCCCCCTCTTAAATTTGTTGATTAAGAGGCGAGCTGCT 1560  
1501 AGAAAGAAATGTTGCCCCCCTCTTAAATTTGTTGATTAAGAGGCGAGCTGCT 1560  
1561 ACTCCATTCAGATCTGCTGATGATGAGAACAGTATGAGACTTGAAGAGTGAAC 1620  
1561 ACTCCATTCAGATCTGCTGATGATGAGAACAGTATGAGACTTGAAGAGTGAAC 1620  
1621 AAAAAAGTAAAGCCAACTGTGGAGAGAGCAACAGCTGCAAAAGATACAGACAA 1680  
1621 AAAAAAGTAAAGCCAACTGTGGAGAGAGCAACAGCTGCAAAAGATACAGACAA 1680  
1681 CTGCGATGATCTTCTGCGACAGAAACATGATGATGATGATGATGATGATGAT 1740  
1681 CTGCGATGATCTTCTGCGACAGAAACATGATGATGATGATGATGATGATGAT 1740  
1741 ATGACAGCATAGTATGATTAAGGAGAGAGCAAAAGCAGATCCCAATGCTGCTGCTG 1800  
1741 ATGACAGCATAGTATGATTAAGGAGAGAGCAAAAGCAGATCCCAATGCTGCTGCTG 1800  
1801 GAACCTTATCATCATTAATCCGATTTATTTGACATGATGATGATGATGATGAT 1860  
1801 GAACCTTATCATCATTAATCCGATTTATTTGACATGATGATGATGATGATGAT 1860  
1861 GTGCTACTCTGCTGCTGCGAGATCAAGTCTAAAGGTTTAAAGCTGATCAAGAACCC 1920  
1861 GTGCTACTCTGCTGCTGCGAGATCAAGTCTAAAGGTTTAAAGCTGATCAAGAACCC 1920  
1921 GCTTATCATGATGATGATTAAGGAGAGCACTACAGCAAGAGATGATGATGATGAT 1980  
1921 GCTTATCATGATGATGATTAAGGAGAGCACTACAGCAAGAGATGATGATGATGAT 1980  
1981 AGTTGCTGTTACTGCCACCCCGATCAATATTCATATGATGATGATGATGATGAT 2040  
1981 AGTTGCTGTTACTGCCACCCCGATCAATATTCATATGATGATGATGATGATGAT 2040  
2041 AGGAGATATCAAAAGAGAGTCTCAGATCAAGCTTCCAACTTGCAGACAGAAATAC 2100  
2041 AGGAGATATCAAAAGAGAGTCTCAGATCAAGCTTCCAACTTGCAGACAGAAATAC 2100  
2101 AACGAGGGAAGATATCTGAAAAACAATGAAAGAGAACTTTGTTGATGATG 2160  
2101 AACGAGGGAAGATATCTGAAAAACAATGAAAGAGAACTTTGTTGATGATGATG 2160  
2161 AAGATGCTCAGAGTGTGTAAGAGAGTCAAAATTTCCCTCTCAGCTCAGACAGCTG 2220  
2161 AAGATGCTCAGAGTGTGTAAGAGAGTCAAAATTTCCCTCTCAGCTCAGACAGCTG 2220  
2221 AAGTTGCAACAGAGCTCTATGCAAAATTTTGAACCTTTGCAACATTTGAACCT 2280  
2221 AAGTTGCAACAGAGCTCTATGCAAAATTTTGAACCTTTGCAACATTTGAACCT 2280  
2281 AATATATGATGATTTATTTAACTCAGATCAAAACAGCTGTGCGCAACCTGAAAGAT 2340  
2281 AATATATGATGATTTATTTAACTCAGATCAAAACAGCTGTGCGCAACCTGAAAGAT 2340  
2341 TTGAGAGATCATTTAACCAGAAACATTTTGGGTGATGATGATGATGATGATGAT 2400  
2341 TTGAGAGATCATTTAACCAGAAACATTTTGGGTGATGATGATGATGATGATGAT 2400

|||||  
Db 2341 TTGAAGAGCATTAACCGAGAAACATTTGGGTAGCATCTGAATCTCAGAGAAACAA 2400  
QY 2401 ACCAGCTGAAGAGATGAGAGATCATTTAGCATTTTCATCAAGATAGACATGCACTGTAAGG 2460  
Db 2401 ACCAGCTGAAGAGATGAGAGATCATTTAGCATTTTCATCAAGATAGACATGCACTGTAAGG 2460  
QY 2461 AATGCAAGAAATTTAACTCAATGTTTGCAATCATCATGAGTGGCCCTAAACCTGGACCAAGTGG 2520  
Db 2461 AATGCAAGAAATTTAACTCAATGTTTGCAATCATCATGAGTGGCCCTAAACCTGGACCAAGTGG 2520  
QY 2521 CAAGACTGGCAACGACCTGGGAGAAACCTCCCAATTAATACGAAAACTATTTCAAGATC 2580  
Db 2521 CAAGACTGGCAACGACCTGGGAGAAACCTCCCAATTAATACGAAAACTATTTCAAGATC 2580  
QY 2581 TCCAAACACCTGTTGATCTCTCCAGAAACATGCGAAATATGCTAATGTTCTCAATAGTC 2640  
Db 2581 TCCAAACACCTGTTGATCTCTCCAGAAACATGCGAAATATGCTAATGTTCTCAATAGTC 2640  
QY 2641 AAATCTACAACTCCCATTAATCCCTCTATTTCCAGTTATCAAAAAAGATCTCACCTTCC 2700  
Db 2641 AAATCTACAACTCCCATTAATCCCTCTATTTCCAGTTATCAAAAAAGATCTCACCTTCC 2700  
QY 2701 TTCACGAAGAAATGACTCAAAAAGTAGACGGCTGCTCAATTTTGAAGACTAAGATGA 2760  
Db 2701 TTCACGAAGAAATGACTCAAAAAGTAGACGGCTGCTCAATTTTGAAGACTAAGATGA 2760  
QY 2761 TTGCAAAAAGAAATTTGCTCAGCTGGCCGAATGGCTTCACATGGAACATGCACTGCTCTCA 2820  
Db 2761 TTGCAAAAAGAAATTTGCTCAGCTGGCCGAATGGCTTCACATGGAACATGCACTGCTCTCA 2820  
QY 2821 TGTTCAGAGACTCGGAAGAAATGGGGAGTTGGGGTCTCTCAGCAGAGTAGTACAA 2880  
Db 2821 TGTTCAGAGACTCGGAAGAAATGGGGAGTTGGGGTCTCTCAGCAGAGTAGTACAA 2880  
QY 2881 ATGCAACAGTGTAGATGTTGCTCAGACAGAGTGTATAAAAAGCGGGTACGTAGTT 2940  
Db 2881 ATGCAACAGTGTAGATGTTGCTCAGACAGAGTGTATAAAAAGCGGGTACGTAGTT 2940  
QY 2941 CCTTTTCATATGCCAAAAAGCTTTATGAGATATGCCCAATGGCTCGAAAAAGTACAGACT 3000  
Db 2941 CCTTTTCATATGCCAAAAAGCTTTATGAGATATGCCCAATGGCTCGAAAAAGTACAGACT 3000  
QY 3001 ACCTTTCAATTTGAGACTGGAATGAGACGAGAGAGTCTTCAGACATTTTCTCTGAGT 3060  
Db 3001 ACCTTTCAATTTGAGACTGGAATGAGACGAGAGAGTCTTCAGACATTTTCTCTGAGT 3060  
QY 3061 GTGAGCCAGCAACACACATTTGCTTAAGATCTGTGTGACAAAAAGCCTGTCAAAATCCG 3120  
Db 3061 GTGAGCCAGCAACACACATTTGCTTAAGATCTGTGTGACAAAAAGCCTGTCAAAATCCG 3120  
QY 3121 AGACCTCTCCAGTAGTCCCAAGGGGAGGGTCAACAAGAAAGTCCAGTCCCTGCCACAGC 3180  
Db 3121 AGACCTCTCCAGTAGTCCCAAGGGGAGGGTCAACAAGAAAGTCCAGTCCCTGCCACAGC 3180  
QY 3181 CCCAGAGACGACCAACACGACATTAATCAACAGAGGATACAGGTTCCGCCGCTGT 3240  
Db 3181 CCCAGAGAGCAGCAACCAACGACATTAATCAACAGAGGATACAGGTTCCGCCGCTGT 3240  
QY 3241 CCTTTATCTTCACGAGAAAGTGGCCGTAAAGATCTCCACCTTTTGGCATAACT 3300  
Db 3241 CCTTTATCTTCACGAGAAAGTGGCCGTAAAGATCTCCACCTTTTGGCATAACT 3300  
QY 3301 CTCACACAGCTTTAAAAAAATCTTTCTTGTCTGAAGAGAGAGTTTGAAGCTCAC 3360  
Db 3301 CTCACACAGCTTTAAAAAAATCTTTCTTGTCTGAAGAGAGAGTTTGAAGCTCAC 3360  
QY 3361 AGAAGACGCTGAAGATACATATCAAAATGCATCTTCGAGCTTTTCTCTCTCTACTT 3420  
Db 3361 AGAAGACGCTGAAGATACATATCAAAATGCATCTTCGAGCTTTTCTCTCTCTACTT 3420  
QY 3421 CTCACAGAGTCTTCCAGAGAAAGGCTATCTTGGCTCCAGTGTACTGTGGATATTT 3480  
|||||

Db 3421 CTCACAGAGTCTTCCAGAGAAAGGCTATCTTGGCTCCAGTGTACTGTGGATATTT 3480  
QY 3481 TTTTCACATTTCTGTCTACAGTGAATTTTCTTACAGATCCAGTATTTGTTAGCAATTCGCTT 3540  
Db 3481 TTTTCACATTTCTGTCTACAGTGAATTTTCTTACAGATCCAGTATTTGTTAGCAATTCGCTT 3540  
QY 3541 TTGACTCAGTGGCAGTCTCACTGCAGATGAGAGGGCCGAGAGCAATTTGTCAGATG 3600  
Db 3541 TTGACTCAGTGGCAGTCTCACTGCAGATGAGAGGGCCGAGAGCAATTTGTCAGATG 3600  
QY 3601 TGGAAACAAACCTTAGGAGATGGGAGAGTGGAGGGGAGACCATGATTGAACCTGATCAGT 3660  
Db 3601 TGGAAACAAACCTTAGGAGATGGGAGAGTGGAGGGGAGACCATGATTGAACCTGATCAGT 3660  
QY 3661 ATAGCTTGGGGTCTTATGACCAATGTCCGAGGGCCGAGGCTTATATGCTACAGCTACAG 3720  
Db 3661 ATAGCTTGGGGTCTTATGACCAATGTCCGAGGGCCGAGGCTTATATGCTACAGCTACAG 3720  
QY 3721 TAATTTCTTCTCCACAGCACAGAGCACTTTCCAGGATCAGGGGGATCGCGCTCACCTTG 3780  
Db 3721 TAATTTCTTCTCCACAGCACAGAGCACTTTCCAGGATCAGGGGGATCGCGCTCACCTTG 3780  
QY 3781 ATGCTGCTGACAGTGGCCGTGGAGGCTGAGACGTATGCTCAAGTGGCTCCCATGATATA 3840  
Db 3781 ATGCTGCTGACAGTGGCCGTGGAGGCTGAGACGTATGCTCAAGTGGCTCCCATGATATA 3840  
QY 3841 TACAGACGATCCAGCACACAGAGAAAGCTGGAGACTCTTCCATTGCGGCAATCTCTTG 3900  
Db 3841 TACAGACGATCCAGCACACAGAGAAAGCTGGAGACTCTTCCATTGCGGCAATCTCTTG 3900  
QY 3901 ATTATTCAGGGAGTCTGCGAGGTTTATGGCATCAAGACCCATATGAGCAAAATATGT 3960  
Db 3901 ATTATTCAGGGAGTCTGCGAGGTTTATGGCATCAAGACCCATATGAGCAAAATATGT 3960  
QY 3961 TTTCTGATCATACACAAAGTATACAGGCAAAATACAAAGTAGAGAGGCTTGAACAAAG 4020  
Db 3961 TTTCTGATCATACACAAAGTATACAGGCAAAATACAAAGTAGAGAGGCTTGAACAAAG 4020  
QY 4021 CCCAGTCCGAGCAAGCTGGGCGTCTTCCACAGGTTACTGTGGGAGAAAGACTCGAAAGGTG 4080  
Db 4021 CCCAGTCCGAGCAAGCTGGGCGTCTTCCACAGGTTACTGTGGGAGAAAGACTCGAAAGGTG 4080  
QY 4081 ACACAGGCACAATTAAGCGGAGGGGTGAAGAGATGTTTCATTGAAACCGGAAGCAGTA 4140  
Db 4081 ACACAGGCACAATTAAGCGGAGGGGTGAAGAGATGTTTCATTGAAACCGGAAGCAGTA 4140  
QY 4141 GCCTAAAGTCTGTGACTACGGAAGAAACCAAGCTGTGCCATGCCCTGCACATAGCTG 4200  
Db 4141 GCCTAAAGTCTGTGACTACGGAAGAAACCAAGCTGTGCCATGCCCTGCACATAGCTG 4200  
QY 4201 TGGCATCAAGTACTACAAAGGGGCTATTGCAAGAAAGAGGGCAGGATTCGAGAGCCCC 4260  
Db 4201 TGGCATCAAGTACTACAAAGGGGCTATTGCAAGAAAGAGGGCAGGATTCGAGAGCCCC 4260  
QY 4261 CGCCCAACCCCTCCGGCTATCTGGAATTTCCATTAATCTGACTTTCCAGAAAGGACACTCC 4320  
Db 4261 CGCCCAACCCCTCCGGCTATCTGGAATTTCCATTAATCTGACTTTCCAGAAAGGACACTCC 4320  
QY 4321 ATCCAGCCAGAAACCGCCGACTACAAAGTGGCCCTTCAGAGATCGCGGATGTCGCAC 4380  
Db 4321 ATCCAGCCAGAAACCGCCGACTACAAAGTGGCCCTTCAGAGATCGCGGATGTCGCAC 4380  
QY 4381 GATTCCTCCGACACAGCTGGGCTTTCATCCGTACAGCAGCCACATGGGATCCACACAGCA 4440  
Db 4381 GATTCCTCCGACACAGCTGGGCTTTCATCCGTACAGCAGCCACATGGGATCCACACAGCA 4440  
QY 4441 GCAGGCTGTGAACAAACCTCAGTGGCATTAACCGAAGCGATGCCGCTGGCC 4500  
Db 4441 GCAGGCTGTGAACAAACCTCAGTGGCATTAACCGAAGCGATGCCGCTGGCC 4500  
QY 4501 CTTATAGTCCCAAGGATTTTCCACGAGAGAGATGAAGATGAACAAGTTTCTGCTGTTT 4560  
Db 4501 CTTATAGTCCCAAGGATTTTCCACGAGAGAGATGAAGATGAACAAGTTTCTGCTGTTT 4560

QY 4561 GAGGCACAGACTTTCTGGAAGCAGAGCCACCCTGAAGAAGAGACACAAGACGT 4620  
|||||  
Db 4561 GAGGCACAGACTTTCTGGAAGCAGAGCCACCCTGAAGAAGAGAGACACAAGACGT 4620  
QY 4621 CCTGAGCATTTGGAGCTTTGGAACATCTGAGAGCGGTGAGACAGTTGGCTCTTC 4680  
|||||  
Db 4621 CCTGAGCATTTGGAGCTTTGGAACATCTGAGAGCGGTGAGACAGTTGGCTCTTC 4680  
QY 4681 CCTGCTTTAAAGACAGATGGGGCTTCTGCCCCCTCTTCCTTTGCCCTTGATGTA 4740  
|||||  
Db 4681 CCTGCTTTAAAGACAGATGGGGCTTCTGCCCCCTCTTCCTTTGCCCTTGATGTA 4740  
QY 4741 AATCTGTGAAGAAATGGCCCTGGCAGCTTTTCAGACTTTGTTCCTTGAATGACAGCTGC 4800  
|||||  
Db 4741 AATCTGTGAAGAAATGGCCCTGGCAGCTTTTCAGACTTTGTTCCTTGAATGACAGCTGC 4800  
QY 4801 AGCAATCTTGCAGCTCCCACTGTTGCTGCTGCGACATCACACAGTATCATTCCAAATTC 4860  
|||||  
Db 4801 AGCAATCTTGCAGCTCCCACTGTTGCTGCTGCGACATCACACAGTATCATTCCAAATTC 4860  
QY 4861 CAAGATCTATCAACACAGATGATTCACCTGGCTGCACTTCTCATGCTTGGAAGATTT 4920  
|||||  
Db 4861 CAAGATCTATCAACACAGATGATTCACCTGGCTGCACTTCTCATGCTTGGAAGATTT 4920  
QY 4921 TTTTAAATCTCTTAAATTTAGATTTCAATCCAGTCTAGCAGTTGATCTGATGGATATG 4980  
|||||  
Db 4921 TTTTAAATCTCTTAAATTTAGATTTCAATCCAGTCTAGCAGTTGATCTGATGGATATG 4980  
QY 4981 AAAAAAGCAGCATTTGAATCTTGGGGCCCTTAAACCACAGAGAGACAAAGAAAA 5040  
|||||  
Db 4981 AAAAAAGCAGCATTTGAATCTTGGGGCCCTTAAACCACAGAGAGACAAAGAAAA 5040  
QY 5041 CAATGAATCTCTTGAATGACAGTCTTGTCACTGTTTACAGTCTCTCTTTAAAA 5100  
|||||  
Db 5041 CAATGAATCTCTTGAATGACAGTCTTGTCACTGTTTACAGTCTCTCTTTAAAA 5100  
QY 5101 AAAAAATGCTTTAAAGATTTTGTTCAGAGAGTAATATATATCATTTAATGATTAAG 5160  
|||||  
Db 5101 AAAAAATGCTTTAAAGATTTTGTTCAGAGAGTAATATATATATCATTTAATGATTAAG 5160  
QY 5161 TTTTAAATTTAAACCTTAAGTAGGGTTGCCAGCCTGGTTCTGAAAAACCAATATGCCGG 5220  
|||||  
Db 5161 TTTTAAATTTAAACCTTAAGTAGGGTTGCCAGCCTGGTTCTGAAAAACCAATATGCCGG 5220  
QY 5221 ACAGGGTGTGGCCACACCAAGAGAGCGGAAGACTGGCTTGTGACCTGGCTTCCATG 5280  
|||||  
Db 5221 ACAGGGTGTGGCCACACCAAGAGAGCGGAAGACTGGCTTGTGACCTGGCTTCCATG 5280  
QY 5281 TCCCTGTGGCTCAACCGGAGTGGCTATCTGGAATGATGAATGTTAGCCAAATTA 5340  
|||||  
Db 5281 TCCCTGTGGCTCAACCGGAGTGGCTATCTGGAATGATGAATGTTAGCCAAATTA 5340  
QY 5341 TACCAAGACACCTATCTGCTCTTCCCAAGTGAATGGGGTCTTGTAAAACTGTTTG 5400  
|||||  
Db 5341 TACCAAGACACCTATCTGCTCTTCCCAAGTGAATGGGGTCTTGTAAAACTGTTTG 5400  
QY 5401 CACATGGCCAGGGGAGGAGAACTAGAACCTTGTGCTCTGAGGCTTATGAGGACAG 5460  
|||||  
Db 5401 CACATGGCCAGGGGAGGAGAACTAGAACCTTGTGCTCTGAGGCTTATGAGGACAG 5460  
QY 5401 CACATGGCCAGGGGAGGAGAACTAGAACCTTGTGCTCTGAGGCTTATGAGGACAG 5460  
|||||  
Db 5401 CACATGGCCAGGGGAGGAGAACTAGAACCTTGTGCTCTGAGGCTTATGAGGACAG 5460  
QY 5461 ACAGTGTATGGCGAGTGTCTCTCCATTCAGATGAGATGGAACCCCAATTTTAA 5520  
|||||  
Db 5461 ACAGTGTATGGCGAGTGTCTCTCCATTCAGATGAGATGGAACCCCAATTTTAA 5520  
QY 5521 GTTATATTTCTTTGATTTTGTAAATTTAGAGTGAAGTGTGTTTTTTTGT 5580  
|||||  
Db 5521 GTTATATTTCTTTGATTTTGTAAATTTAGAGTGAAGTGTGTTTTTTTGT 5580  
QY 5581 TTTTAAAGAGAACTTTATATACGATGAGATGCAATGGAAGAGAGCTTGGATGTT 5640  
|||||  
Db 5581 TTTTAAAGAGAACTTTATATACGATGAGATGCAATGGAAGAGAGCTTGGATGTT 5640

QY 5641 GGAGCTAATGCCAGTGTGTTATACGTCTTTCAAGACAGCTCCCTTATTTGATTTGCC 5700  
|||||  
Db 5641 GGAGCTAATGCCAGTGTGTTATACGTCTTTCAAGACAGCTCCCTTATTTGATTTGCC 5700  
QY 5701 ATTAGGAATTAACAAGCTTTTAAACGTGATTAAGATCAAAAACCTGTTAGCATGCC 5760  
|||||  
Db 5701 ATTAGGAATTAACAAGCTTTTAAACGTGATTAAGATCAAAAACCTGTTAGCATGCC 5760  
QY 5761 AGCTTTTGCAGGCGAGTTAGTACACCAAGACTAACCTTCAAGGGGCTTTATGAGCGCTG 5820  
|||||  
Db 5761 AGCTTTTGCAGGCGAGTTAGTACACCAAGACTAACCTTCAAGGGGCTTTATGAGCGCTG 5820  
QY 5821 CATATAGAGAAGCCCTAAGTGTAGCAACACTGCTCAGCTGCTATTAACCTATTAAT 5880  
|||||  
Db 5821 CATATAGAGAAGCCCTAAGTGTAGCAACACTGCTCAGCTGCTATTAACCTATTAAT 5880  
QY 5881 GACTGAATGACCCCTCAGCTATTTTGTGTTGTTTGCACAGACTCCGGAAAAAGTA 5940  
|||||  
Db 5881 GACTGAATGACCCCTCAGCTATTTTGTGTTGTTTGCACAGACTCCGGAAAAAGTA 5940  
QY 5941 AGCTGCCAATGCGAGTACTACTCAAAATGTGAGAACTGCTGCTTGATTTTTC 6000  
|||||  
Db 5941 AGCTGCCAATGCGAGTACTACTCAAAATGTGAGAACTGCTGCTTGATTTTTC 6000  
QY 6001 ATTAATTCAGCTGATCATATTTGATCAGTAGATTAACGTAATAGCTTCAAAATTTTAA 6060  
|||||  
Db 6001 ATTAATTCAGCTGATCATATTTGATCAGTAGATTAACGTAATAGCTTCAAAATTTTAA 6060  
QY 6061 GTGGAATGTAGGTTTCTTCTGCTGATCAACAAATGTGAGGCTTTATTAATATCT 6120  
|||||  
Db 6061 GTGGAATGTAGGTTTCTTCTGCTGATCAACAAATGTGAGGCTTTATTAATATCT 6120  
QY 6121 CTCTGATCATGCAATTTGCTACTTCTGCTTATTAATGTCATTTGATTTGATTT 6180  
|||||  
Db 6121 CTCTGATCATGCAATTTGCTACTTCTGCTTATTAATGTCATTTGATTTGATTT 6180  
QY 6181 TTACATGTAATGATCATTTATTTGCCAGTTTATTAATAGCTTATGAGCTTATGAT 6240  
|||||  
Db 6181 TTACATGTAATGATCATTTATTTGCCAGTTTATTAATAGCTTATGAGCTTATGAT 6240  
QY 6241 ATGAAGAAGCAAAATGTACTCTACACAAAGTGCACAAATGTTATCTAAGCATTAAGT 6300  
|||||  
Db 6241 ATGAAGAAGCAAAATGTACTCTACACAAAGTGCACAAATGTTATCTAAGCATTAAGT 6300  
QY 6301 AATGTAGAACATAGACCTGATCTAGTCTGCTGATGTAAGTGCAGAAATGTA 6360  
|||||  
Db 6301 AATGTAGAACATAGACCTGATCTAGTCTGCTGATGTAAGTGCAGAAATGTA 6360  
QY 6361 CAATTAACCTGCTGATTTTCCATACCTTTGATACCTTGTACCTGATGCTTTTAA 6420  
|||||  
Db 6361 CAATTAACCTGCTGATTTTCCATACCTTTGATACCTTGTACCTGATGCTTTTAA 6420  
QY 6421 AGACATTTGCTGATCTGATACCTTTTGTATTTTAAATTAATTTGATTAATTTGGT 6480  
|||||  
Db 6421 AGACATTTGCTGATCTGATACCTTTTGTATTTTAAATTAATTTGATTAATTTGGT 6480  
QY 6481 TATATTTTGTGAAGATGTAGAAATGTACTATGTTTATGCTTCAATCAGAGTTTGA 6540  
|||||  
Db 6481 TATATTTTGTGAAGATGTAGAAATGTACTATGTTTATGCTTCAATCAGAGTTTGA 6540  
QY 6541 CAAGCTGGAATAATTAATTAATTAACAT 6568  
|||||  
Db 6541 CAAGCTGGAATAATTAATTAATTAACAT 6568

## RESULT 2

US-09-911-826a-1  
; Sequence 1, Application US/09911826a

; Patent No. US20020143164A1

## ; GENERAL INFORMATION:

; APPLICANT: Rolin, Daniela and Pham, Nam  
; TITLE OF INVENTION: RAS Activator Nucleic Acid Molecules, Polypeptides and  
; TITLE OF INVENTION: Methods of Use





```
QY 1861 GTCCTACTCCTGACTTGCAGATCAAGTCTAAGGGTTTTTAAAGGCTGATCAGCAAAAGCC 1920
    |||
Db 1861 GTCTTACTCTGACTTGCAGATCAAGTCTAAGGGTTTTTAAAGGCTGATCAGCAAAAGCC 1920
QY 1921 GCTACATCATGATCATAGTAGAGACATCAGCAAGAAAGAAAGTGGTCAATTCAGGGCATCAGGG 1980
    |||
Db 1921 GCTACATCATGATCATAGTAGAGACATCAGCAAGAAAGAAAGTGGTCAATTCAGGGCATCAGGG 1980
QY 1981 AGTTTGGTGTACTGGCACCCCGGATCAATATTCACATATGTAGAGTCTGTGTACACCTG 2040
    |||
Db 1981 AGTTTGGTGTACTGGCACCCCGGATCAATATTCACATATGTAGAGTCTGTGTACACCTG 2040
QY 2041 AGGAGATTAATCAAAACAAGAAAGCTTCAGATCAGCTTTCACAACTTTCAGACAGAAATAC 2100
    |||
Db 2041 AGGAGATTAATCAAAACAAGAAAGCTTCAGATCAGCTTTCACAACTTTCAGACAGAAATAC 2100
QY 2101 AACTGAGTGGAAAGTATTATCTGAAAAACAACATGGAAACAGAAAGTCTTGTTCAGATG 2160
    |||
Db 2101 AACTGAGTGGAAAGTATTATCTGAAAAACAACATGGAAACAGAAAGTCTTGTTCAGATG 2160
QY 2161 AAGATGCTCAGAGAGTGTGTGAGAGAGTCAAAATTTCCCTCTCAGCTCAGCAGCTGTG 2220
    |||
Db 2161 AAGATGCTCAGAGAGTGTGTGAGAGAGTCAAAATTTCCCTCTCAGCTCAGCAGCTGTG 2220
QY 2221 AAGTTGCAACACAGCTCTCTATGCGAAATTTTGAACCTTTTCGACATTTGAACCTTCTG 2280
    |||
Db 2221 AAGTTGCAACACAGCTCTCTATGCGAAATTTTGAACCTTTTCGACATTTGAACCTTCTG 2280
QY 2281 AATATATGATGATTTTATTTAACTCAGATCAAAACAGCTGTGCAACCTGAGAGAT 2340
    |||
Db 2281 AATATATGATGATTTTATTTAACTCAGATCAAAACAGCTGTGCAACCTGAGAGAT 2340
QY 2341 TTGAAGAAGTCAATTAACAGAAACATTTTGGGTAGCATCTGAAATTTCTCGAAGAAACAA 2400
    |||
Db 2341 TTGAAGAAGTCAATTAACAGAAACATTTTGGGTAGCATCTGAAATTTCTCGAAGAAACAA 2400
QY 2401 ACCACTGAAGAGATGAAGATCATTAGCATTTTCATCAAGATAGCATGCACTGTAGGG 2460
    |||
Db 2401 ACCACTGAAGAGATGAAGATCATTAGCATTTTCATCAAGATAGCATGCACTGTAGGG 2460
QY 2461 AATGCAAGAAATTTTAACTCAATGTTTGCATATCAGTGGCTTAACCTGGCAGCAGTGG 2520
    |||
Db 2461 AATGCAAGAAATTTTAACTCAATGTTTGCATATCAGTGGCTTAACCTGGCAGCAGTGG 2520
QY 2521 CAAGACTGGGAAGCAGCTGGGGAACCTCCCAATTAATAGCAAAACTATTTCAGATC 2580
    |||
Db 2521 CAAGACTGGGAAGCAGCTGGGGAACCTCCCAATTAATAGCAAAACTATTTCAGATC 2580
QY 2581 TTCGAAGACCTGTTTGAATCTTCCAGAAACATGGCAAAATATCGTAATGTTCTCAATAGTC 2640
    |||
Db 2581 TTCGAAGACCTGTTTGAATCTTCCAGAAACATGGCAAAATATCGTAATGTTCTCAATAGTC 2640
QY 2641 AAAATCTAACCTCCCATTAATCCCTCTATTTCCAGTTATCAAAAAGATCTCACCTTCC 2700
    |||
Db 2641 AAAATCTAACCTCCCATTAATCCCTCTATTTCCAGTTATCAAAAAGATCTCACCTTCC 2700
QY 2701 TTCAGCAAGAAATGACTCAAAAGTAGAGGGGTGCTCAATTTTGAAGAGTAAAGATGA 2760
    |||
Db 2701 TTCAGCAAGAAATGACTCAAAAGTAGAGGGGTGCTCAATTTTGAAGAGTAAAGATGA 2760
QY 2761 TTGCAAAAGAAATTCGTCAGAGTTGGCCGAAATGGCTTCACTGAACATGAGCCCTGCTTA 2820
    |||
Db 2761 TTGCAAAAGAAATTCGTCAGAGTTGGCCGAAATGGCTTCACTGAACATGAGCCCTGCTTA 2820
QY 2821 TGTTCAGAGACTCGGAAGAAAGAAATGGCGAGTTTGGGGTCTCTCAGCCAGGGTAGTACA 2880
    |||
Db 2821 TGTTCAGAGACTCGGAAGAAAGAAATGGCGAGTTTGGGGTCTCTCAGCCAGGGTAGTACA 2880
QY 2881 ATGCAGACAGTGTAGATGTTGTCTCAGACAGGTGTATAAAAAGCGGGTACGTCTGATTT 2940
    |||
Db 2881 ATGCAGACAGTGTAGATGTTGTCTCAGACAGGTGTATAAAAAGCGGGTACGTCTGATTT 2940
QY 2941 CTTTCTCAATGCCAAAAGCTTTATGAGATGCCCAATGGCTCGAAAAAGTGAACAGT 3000
    |||

Db 2941 CTTTCTCAATGCCAAAAGCTTTATGAGATGCCCAATGGCTCGAAAAAGTGAACAGT 3000
    |||
QY 3001 ACCTTTCATTTGAGCTAGAAATGACAGAGAGATCTTTCAGACATTTTCTGTAGT 3060
    |||
Db 3001 ACCTTTCATTTGAGCTAGAAATGACAGAGAGATCTTTCAGACATTTTCTGTAGT 3060
QY 3061 GTGAGCCAGCAACCAACATTTGGCTTAAGAAATCTGTGACAAAAAGCCTGTCAATCCG 3120
    |||
Db 3061 GTGAGCCAGCAACCAACATTTGGCTTAAGAAATCTGTGACAAAAAGCCTGTCAATCCG 3120
QY 3121 AGACCTTCTCAGTAGCTTCCAAAGGCGAGGTGACAAACAGAAAGTCACTCCCTGCAAGC 3180
    |||
Db 3121 AGACCTTCTCAGTAGCTTCCAAAGGCGAGGTGACAAACAGAAAGTCACTCCCTGCAAGC 3180
QY 3181 CCCAGCAGCAGCCACCACAGACATTAATTAACACAGAGACTCAGATTCGCCCGCTGT 3240
    |||
Db 3181 CCCAGCAGCAGCCACCACAGACATTAATTAACACAGAGACTCAGATTCGCCCGCTGT 3240
QY 3241 CCGTTATCCCTCAGGAAAGAAAGTCCCGTAAAGATCTCCACCTTTTGGCATAAACT 3300
    |||
Db 3241 CCGTTATCCCTCAGGAAAGAAAGTCCCGTAAAGATCTCCACCTTTTGGCATAAACT 3300
QY 3301 CTCACACAGCTTTAAAAAAATTTCTTGTCTGAAAGAAAGTGTGGAAGCTCACCA 3360
    |||
Db 3301 CTCACACAGCTTTAAAAAAATTTCTTGTCTGAAAGAAAGTGTGGAAGCTCACCA 3360
QY 3361 AGAAAGAGCTGAAGTATCAATATCAAAATGATCTTCGAGCTTCTCTCCTACTT 3420
    |||
Db 3361 AGAAAGAGCTGAAGTATCAATATCAAAATGATCTTCGAGCTTCTCTCCTACTT 3420
QY 3421 CTCACAGAGTTCTCCAGAAAGAGCTATATCTTGGCTCCAGTGTACTGTGATATT 3480
    |||
Db 3421 CTCACAGAGTTCTCCAGAAAGAGCTATATCTTGGCTCCAGTGTACTGTGATATT 3480
QY 3481 TTTCAGATTCGTGTACAGAGAAATTTCTTCAGATCCAGTATGTTTGAACAATTCGTCT 3540
    |||
Db 3481 TTTCAGATTCGTGTACAGAGAAATTTCTTCAGATCCAGTATGTTTGAACAATTCGTCT 3540
QY 3541 TTGACTCAGTGCAGTCTCACTGACAGATGAGAGGCGCAGAGCTTCTGTACATCG 3600
    |||
Db 3541 TTGACTCAGTGCAGTCTCACTGACAGATGAGAGGCGCAGAGCTTCTGTACATCG 3600
QY 3601 TCGAACAACACCTTAGGATGCGAGATGAGAGGCGGACATGATTTGAACCTGATCAGT 3660
    |||
Db 3601 TCGAACAACACCTTAGGATGCGAGATGAGAGGCGGACATGATTTGAACCTGATCAGT 3660
QY 3661 ATACTTGGGGTCCCTATGACCAATGTCGAGGCGGAGGCTTATATGCTACAGCTACAG 3720
    |||
Db 3661 ATACTTGGGGTCCCTATGACCAATGTCGAGGCGGAGGCTTATATGCTACAGCTACAG 3720
QY 3721 TAAATTTCTTCCAAAGCAGAGAGAACTTTCCAGATCAGGGGATCCCGGCTCACTTG 3780
    |||
Db 3721 TAAATTTCTTCCAAAGCAGAGAGAACTTTCCAGATCAGGGGATCCCGGCTCACTTG 3780
QY 3781 ATGCTGTGACAGTGGGCGGAGGAGTGGACGTATGCTCAATGCTCCCATGATATA 3840
    |||
Db 3781 ATGCTGTGACAGTGGGCGGAGGAGTGGACGTATGCTCAATGCTCCCATGATATA 3840
QY 3841 TACAGAGATTCACAGCAGAGAGAGTGGAGACTCTTCAATTTGGGCACTACTGCTTG 3900
    |||
Db 3841 TACAGAGATTCACAGCAGAGAGAGTGGAGACTCTTCAATTTGGGCACTACTGCTTG 3900
QY 3901 ATTATTTCAAGGGATCTGACAGTTTATAGGCAATCAACAGCATATGACCAATTAATGT 3960
    |||
Db 3901 ATTATTTCAAGGGATCTGACAGTTTATAGGCAATCAACAGCATATGACCAATTAATGT 3960
QY 3961 TTTCTGATCATAGCACAAGATATAACAGGCAAAATCAAAAGTAGAGAGCCTTGAACAG 4020
    |||
Db 3961 TTTCTGATCATAGCACAAGATATAACAGGCAAAATCAAAAGTAGAGAGCCTTGAACAG 4020
QY 4021 CCCAGTCCGAGCAAGCTGGGCGCTTTCACAGGTTACTGGGGAGAACTCAGAAAGTG 4080
    |||
Db 4021 CCCAGTCCGAGCAAGCTGGGCGCTTTCACAGGTTACTGGGGAGAACTCAGAAAGTG 4080
    |||
```

Db 4021 CCCAGTCCCCGAGCAAGCTGGGCGTCTTCCACAGATTACTGGGGAGAAAGACTCAGAAAGTG 4080  
Qy 4081 ACACAGGCAAAATTAAGCGGAGGGGTGGAAAGATGTTTCCATTGAGCGCGAAAGCAGTA 4140  
Db 4081 ACACAGGCAAAATTAAGCGGAGGGGTGGAAAGATGTTTCCATTGAGCGCGAAAGCAGTA 4140  
Qy 4141 GCCTAACGTCGTGTGATAGGAAAGAAACCAAGCGCTGCCCATGCTGCCCATATAGTGTG 4200  
Db 4141 GCCTAACGTCGTGTGATAGGAAAGAAACCAAGCGCTGCCCATGCTGCCCATATAGTGTG 4200  
Qy 4201 TGGCATCAAGTACTACAAAGGGGCTCATTTGACAGAAAGAGGGGAGGTATCGAGAGCCCC 4260  
Db 4201 TGGCATCAAGTACTACAAAGGGGCTCATTTGACAGAAAGAGGGGAGGTATCGAGAGCCCC 4260  
Qy 4261 CGGCCACCCCTCCGGGCTCATTTGGAATTCCTTACTGACTTTTCCAGAAAGGGGACTCC 4320  
Db 4261 CGGCCACCCCTCCGGGCTCATTTGGAATTCCTTACTGACTTTTCCAGAAAGGGGACTCC 4320  
Qy 4321 ATCCAGCCAGAAACCGCGGACTACAGGTGGCCCTTCAGAGATCGCGATGGTGCAC 4380  
Db 4321 ATCCAGCCAGAAACCGCGGACTACAGGTGGCCCTTCAGAGATCGCGATGGTGCAC 4380  
Qy 4381 GATCTCCGACACAGCTGGGCGCTTCAATCCGTACAGACGCCACATGGGCAATCCACAGCA 4440  
Db 4381 GATCTCCGACACAGCTGGGCGCTTCAATCCGTACAGACGCCACATGGGCAATCCACAGCA 4440  
Qy 4441 GCAGGCTGTGAAACAACCTCAGTGGCATTAACCCGAGAGTGTGACCCGCGCTGCCC 4500  
Db 4441 GCAGGCTGTGAAACAACCTCAGTGGCATTAACCCGAGAGTGTGACCCGCGCTGCCC 4500  
Qy 4501 CTATACGTCCCAAGGGTTTCCACCAGAGAGATGAAGTGAACAAGTTTCTGCTGTT 4560  
Db 4501 CTATACGTCCCAAGGGTTTCCACCAGAGAGATGAAGTGAACAAGTTTCTGCTGTT 4560  
Qy 4561 GAGGCAACAGCTTTTGTGAAGACAGACGACCTGAAAGAGAGACACAGAGACGT 4620  
Db 4561 GAGGCAACAGCTTTTGTGAAGACAGACGACCTGAAAGAGAGACACAAAMAACGT 4620  
Qy 4621 CCGACATTTGGAGCTTGGAACTCATTCTGAGAGCGTGGACAGTTTGGCTCTTC 4680  
Db 4621 CCGACATTTGGAGCTTGGAACTCATTCTGAGAGCGTGGACAGTTTGGCTCTTC 4680  
Qy 4681 CCGCTTTAAAGACAGATGGGGCTTCTCCCTTCTTCCCTTCCCTTTCGATGTGA 4740  
Db 4681 CCGCTTTAAAGACAGATGGGGCTTCTCCCTTCTTCCCTTTCGATGTGA 4740  
Qy 4741 AATACGTGAGAAATTTGCCCTGGCACTTTTTCAGACTTTGTTGCTTGAATGACAGTGC 4800  
Db 4741 AATACGTGAGAAATTTGCCCTGGCACTTTTTCAGACTTTGTTGCTTGAATGACAGTGC 4800  
Qy 4801 AGCAATCTTGAGCTCCCACTGTGTGGCTGCACATCAGACAGATATCAATTCCAAATTC 4860  
Db 4801 AGCAATCTTGAGCTCCCACTGTGTGGCTGCACATCAGACAGATATCAATTCCAAATTC 4860  
Qy 4861 CAAGATCATCAACAAGATGATTCACTGTGCTGCATCTTCAATGCGTGAAGATTT 4920  
Db 4861 CAAGATCATCAACAAGATGATTCACTGTGCTGCATCTTCAATGCGTGAAGATTT 4920  
Qy 4921 TTTTATATCTCTTTTATGATTTCATCCAGTCTTACAGCTGTGATCTCATTTGGGATATG 4980  
Db 4921 TTTTATATCTCTTTTATGATTTCATCCAGTCTTACAGCTGTGATCTCATTTGGGATATG 4980  
Qy 4981 AGAAAGCTAGGCACTGAACTACTTGGGGCTTAAACCAACCAAGAAAGCAAGAAAA 5040  
Db 4981 AGAAAGCTAGGCACTGAACTACTTGGGGCTTAAACCAACCAAGAAAGCAAGAAAA 5040  
Qy 5041 CAATGAATCTTTGAGTACAGTGTGTGCACATTTTACATGTCCTCTTTAAAAA 5100  
Db 5041 CAATGAATCTTTGAGTACAGTGTGTGTGCACATTTTACATGTCCTCTTTAAAAA 5100  
Qy 5101 AAAAAATGATTTAAAGATTTTGTTCAGAGAGTAAATATATATCATTTAATGATTACAG 5160  
Db 5101 AAAAAATGATTTAAAGATTTTGTTCAGAGAGTAAATATATATCATTTAATGATTACAG 5160

Qy 5161 TATTTATTTAAACCTTAAGTAGGGTTGCCAGCCTGGTTCTGAAAAACCAAAATATGCCG 5220  
Db 5161 TATTTATTTAAACCTTAAGTAGGGTTGCCAGCCTGGTTCTGAAAAACCAAAATATGCCG 5220  
Qy 5221 ACAGGCTGTGGCCACACCAAGAGAGGAGACAGTGGCTGTGACCTGGCTTCCATG 5280  
Db 5221 ACAGGCTGTGGCCACACCAAGAGAGGAGACAGTGGCTGTGACCTGGCTTCCATG 5280  
Qy 5281 TCTTCTGTCTTCACCCGGAAGTGCCCTATCTGGAAGTATGAATGTTAGCCAAATTA 5340  
Db 5281 TCTTCTGTCTTCACCCGGAAGTGCCCTATCTGGAAGTATGAATGTTAGCCAAATTA 5340  
Qy 5341 TACCAAGACACCTCATCTCTCTCTCCAGTGGATGGGGTTCTTGTGAAAACTGTTG 5400  
Db 5341 TACCAAGACACCTCATCTCTCTCTCCAGTGGATGGGGTTCTTGTGAAAACTGTTG 5400  
Qy 5401 CACATGGCCAGGGGAGGAGAACTAGAACCTTGTGTCTGTGAGCCTTATGAGAGCAGG 5460  
Db 5401 CACATGGCCAGGGGAGGAGAACTAGAACCTTGTGTCTGTGAGCCTTATGAGAGCAGG 5460  
Qy 5461 ACAGTGTATGGCGGATGTCTCTCATTTGAGATGATGGCAACCCCATTTTAA 5520  
Db 5461 ACAGTGTATGGCGGATGTCTCTCATTTGAGATGATGGCAACCCCATTTTAA 5520  
Qy 5521 GTATATTTCTTTGATTTTGTAAATTTAGAGGTAGGTTTGTGTTTGTGTT 5580  
Db 5521 GTATATTTCTTTGATTTTGTAAATTTAGAGGTAGGTTTGTGTTTGTGTT 5580  
Qy 5581 TTTTAAAGAGAAACATTTATTAAGTGAATGATGATGACATTCAGTGAACAGCTTGGATGT 5640  
Db 5581 TTTTAAAGAGAAACATTTATTAAGTGAATGATGATGACATTCAGTGAACAGCTTGGATGT 5640  
Qy 5641 GAGCTTAATGCCAGCTGTTTATATCTGCTTTTCAAGACAGCTTCCCTTATTTGAATGGC 5700  
Db 5641 GAGCTTAATGCCAGCTGTTTATATCTGCTTTTCAAGACAGCTTCCCTTATTTGAATGGC 5700  
Qy 5701 ATTAGGAAATTAACAAGCCTTTAAGCTGTAAAGTCAAAAAACCGTGTAAACATGCC 5760  
Db 5701 ATTAGGAAATTAACAAGCCTTTAAGCTGTAAAGTCAAAAAACCGTGTAAACATGCC 5760  
Qy 5761 AGCCTTTGCAAGGAGGATTAAGTACCAAAAGACTAACCTCCAGTGTATGAGAGCTG 5820  
Db 5761 AGCCTTTGCAAGGAGGATTAAGTACCAAAAGACTAACCTCCAGTGTATGAGAGCTG 5820  
Qy 5821 CATATAGAGAGGCTTAAGTGTAGCAACATCTGCTCACAGCTGTATTAACCTATAT 5880  
Db 5821 CATATAGAGAGGCTTAAGTGTAGCAACATCTGCTCACAGCTGTATTAACCTATAT 5880  
Qy 5881 GACTGAATGAGCCCTCCACTATTTTGTGTTTGTGTTTGCACAGACTCCGAAAAAGTGA 5940  
Db 5881 GACTGAATGAGCCCTCCACTATTTTGTGTTTGTGTTTGCACAGACTCCGAAAAAGTGA 5940  
Qy 5941 AGGCTCCCAATCTGAGTAGTACTCAAAATGTAGAGAACTGCTGTGTGATTTTTC 6000  
Db 5941 AGGCTCCCAATCTGAGTAGTACTCAAAATGTAGAGAACTGCTGTGTGATTTTTC 6000  
Qy 6001 ATTAATTTACGCTGATCATATTTGATCAGTATGAATTAAGTTCATTAATTTAAAA 6060  
Db 6001 ATTAATTTACGCTGATCATATTTGATCAGTATGAATTAAGTTCATTAATTTAAAA 6060  
Qy 6061 GTGGAATGAGTGTTTTTCACATGATCAAAACAAGTGTGCTGTATTAATTAATTC 6120  
Db 6061 GTGGAATGAGTGTTTTTCACATGATCAAAACAAGTGTGCTGTATTAATTAATTC 6120  
Qy 6121 CTCTGTATCATGGAATTTGTCTACTGTGTTTATTAATTTCAATTAATGATTTGTAAT 6180  
Db 6121 CTCTGTATCATGGAATTTGTCTACTGTGTTTATTAATTTCAATTAATGATTTGTAAT 6180  
Qy 6181 TTACATGTAATATGATTAATTTTCCAGTGTATTAATTAATGAGCTATGAGCTATGTCAT 6240  
Db 6181 TTACATGTAATATGATTAATTTTCCAGTGTATTAATTAATGAGCTATGAGCTATGTCAT 6240



Dh 1897 GGAGTGAACCTGTAGCAAAAGCTGCTGATTCAGACTGAACGTGGTGCATGATTATG 1956  
Qy 1365 GAAGTAAATGGCCAAACTTTGAAAACATTCAGACTGTCAAAACCTATGAAATTCCTAGA 1424  
Dh 1957 GAAGTAAATGGCAAAACTTTGAAAATATTACATTTATGAAACCCGCTTGAAATTTTGAGG 2016  
Qy 1425 AATAACACATCTATCTATCTAGTGAAGAAACATTTATTTATTTAAAGAACTCTA 1484  
Dh 2017 AATAATATCATCTTGTGCACTTACTGTGAAGACCAACATTTTGTGTTCAAGAGTTACTT 2076  
Qy 1485 ACAAGATTGTGAGAGAGAAAGAAATGCTGCCCCACCCTCTTAATTTGGTGACATT 1544  
Dh 2077 TTTAG---GACTGAACAGAGAAATCTGGTGTCTCATATTTCCCAATTTGCTGAAAAA 2133  
Qy 1545 AAAAAGGCCGTGGCTACTCCATTCAGATCTTGCTGTATATGTAAGAACAGTGATAGA 1604  
Dh 2134 AAAAGT---AATGCCATTTCTATCCAGCATGTGCCAGAGATATTTGAAAC---GACTCA 2187  
Qy 1605 CTGAAAAAGTGAACAAAAAGTAAAGCCACAGCTGTG---GGAGAGAGAACAGCTG 1661  
Dh 2188 CAGAGAGAAAGAGTAAAGAAAGTTAAAGCAATATCTTTTCAGGTGAGAGAAACAAATC 2247  
Qy 1662 AAAAGATCTGACAGAACTCGATCAGTATCTTGCCACAGAAACCATCAATGATATT 1721  
Dh 2248 AGGAAGATTGGATAAACACGATTATGATCTGTGCTCCAAAGCTATTATGATGGA 2307  
Qy 1722 GGGATTGGTCACTCAAGATGACAGCATATGATAGATTAAAGCAGAACAGCATCCCA 1781  
Dh 2308 GGGCTAAAGCAATCAGAAATGACAGCATTTGTGGAAACAGGCACTGTAGCATAGCTG 2367  
Qy 1782 ACTGATTTGCTGCTGAGTGAACCTTATCATCCAGTAACTCTGATTTATGAGTGCAT 1841  
Dh 2368 GCTATATATGCCCATCCTGTGAACACATCTCATCCAGACGCCCTGATCTCTGACCTACC 2427  
Qy 1842 CATGCAATTTTAAGCTTCACTGCTACTCTGACTTGCCAGATCAAGTCTAAGGGTTTT 1901  
Dh 2428 ACCAGATTGTGGATTTTTCCAATCTCTTCAGATATCCCTGATCAAGTTATTAAGGTTTC 2487  
Qy 1902 AAGCTGATGAGAAAGCCCTCATCATGATCATGTAAGAGACATCAGCAAAAGAGTG 1961  
Dh 2488 AAGGTGATGAGCAAACTGCTACATTTATCATGATAAAGACACACAGGTAAAGAGTA 2547  
Qy 1962 GTCACTGAGCTATCAGGGAGTTTGTCTGTTACTGACCACCCGATCAATTTTCACTATG 2021  
Dh 2548 GTTTTTCATGCTGTTCATGATTTGGTTGACCGGTGCATCCAGACATATTTCTCTGT 2607  
Qy 2022 GAGCTCTGTCAACCTGAGGAGTAAATCAACAAAGAAAGCTTCCAGATCAGCTTTC 2081  
Dh 2608 GAAGTTTCTGTACTCTGTAGGGTGTCTATAAACAGAGAAAGACTTCCAGATCAGTTCTCC 2667  
Qy 2082 AAATCTGACACAGATATACATGAGTGAAGTATATCTGAAAAACAACATGGAAACA 2141  
Dh 2668 AAATTTACTGATGAATTAACATCAATGAGAGTATTAATAAATAACATGGAACA 2727  
Qy 2142 GAAACTTTTTCAGATGAAGATGTCAGAGTTGTTGAGAGAGATCAATTTCCCTC 2201  
Dh 2728 GAAACCTTATGTTTCAGATGAAGATGCTCAAGAACTAGTTAAGAAACCCAGCTATCCATG 2787  
Qy 2202 CTTCAGCTCAGACTGTGAAGTTGCAACAGAGCTCTATGCGAAATTTTGAACCTTT 2261  
Dh 2788 CTGAGCTGAGTACCATTTGAGTGGCCACAGCTGTCATGAGGGAGCTTTGATTTT 2847  
Qy 2262 CGCAACCTTAACCTTACTGATATATGATGATTTATTTAACTCAGATCAAAACAGC 2321  
Dh 2848 CGAATATTTTGAACCGATGTGATCATGACCTTTTAAAGTTAAATTCACAAACGGA 2907  
Qy 2322 TGTGCCAACCTGAAGATTTGAAGATGTCATTAACAGAGAAACATTTGGTAGCATCT 2381  
Dh 2908 AATATCTATTTGAAGAGTTTGAAGACATTTTAACCAAGAGACATTTCTGGTGGCTCA 2967  
Qy 2382 GAAATTTCTAGAGAAACAAACAGCTGAAGAGATGAAGATCATTTAAGCATTTTCATCAG 2441  
Dh 2968 GAAATTTTAACTGAAGCAATCAGCTCAACGATGAAGATTTATTAAGCATTTTATTA 3027

Qy 2442 ATAGACTGCACTGTAGGGAATGCAAGAAATTTTACTCAATGTTTGCAATCATCAGTGGC 2501  
Dh 3028 ATTGCATCTATTTGTGAGATGATGAACTTCAATTCATGTTTGAATTAATTAAGTGGC 3087  
Qy 2502 CTAAACCTGGCAGCAGTGGCAAGACTGCGAAGCACTGGGAGAAACTTCCCAATTAATAC 2561  
Dh 3088 TTGACCTGGCATCTGTAGACAGATCTGAGAACTGAGAAACTTGGGAAAAAGTTACAGCAATAC 3147  
Qy 2562 GAAAACTATTTCAAGATCTCCAGAGACTGTTGATCTTTTCAGAAACATGGCAAAATAT 2621  
Dh 3148 GAGAAACATCTCAAGATCTACAAAGACATTTTGATCATCTAGAAACATGCGCAAAATAT 3207  
Qy 2622 CGTAATGCTCAATGCTCAAAATCTCAACCTCCCATTAATCCCTATTCCTCAGTTATC 2681  
Dh 3208 AGAATATTTCTTATGATCTCAAGATGACCTCCATTAATTTATTCACCTTCCTGTTGTC 3267  
Qy 2682 AAAAAGATCTCACCTCTCTTCAGAGAGAAATGACCAAAAGTAGAGGGGCTGGTCAAT 2741  
Dh 3268 AAGAAAGATATGACATTTCTACATGAAGAAATGACTCCAAAGATGATGTTTATGTAAC 3327  
Qy 2742 TTTGAGAGCTAAGATGATTTGCAAAAGAAATTCGTACGTTGGCCGAATGGCTTCAGTG 2801  
Dh 3328 TTTGAGAAAGTTAAGATGATTTCCAAAGAAATCCGCAAGTTGTTGAATGACTTGTGCT 3387  
Qy 2802 AACATGACCCCTGCCCATGTTCTGAGACTCGGAAGAAAGAAATGGCGAGTTGGGCTCT 2861  
Dh 3388 AACATGACCCCATGATAGATT-----TCCACAGAGGTCA 3423  
Qy 2862 CTCAGCAGGGTATGTAATATGCAACAGTGTAGATGTTGCTCAGACAGGTGGCTATAA 2921  
Dh 3424 CTGAGTAAAGAAACACAAATTTCAACATCTGATAT---TCAGGAGGTGCTCACAAA 3480  
Qy 2922 AAGCGGTACGTGTGATCTCTTCTCAATGCCAAAAAGCTTTATGAAGTGGCCAAATG 2981  
Dh 3481 AAAAGGCGACGCGCAGCTCTCTCTTAATGCGCAAGAGCTATATGAGATGCCAAATG 3540  
Qy 2982 GCTCGAAATGAAAGCATGATCTTCCATTTGAGCTAAGAAATGACAGAGAGATCTT 3041  
Dh 3541 GCAAGAGAGTGAAGCATATCTTTCCAGTCTGCATGTATGAGACAGATGAGGAGAAAGTTTC 3600  
Qy 3042 CAGCATTTATCTGTCAGTGTGAGCCAGCAACCAACATTTGCTTAAGAAATCTCGGTGAC 3101  
Dh 3601 CAGATGATGATCATTAACATTTGAGGAGACCTGCAATATGATCTTGACCAAGAAATTAAGTAC 3660  
Qy 3102 AAAAAGCTGTCAA---ATCCGAGACTTTCAGTACCTCCAAAGGAGGCTGCACAAAG 3158  
Dh 3661 AAAAGATCAGCAAAATCATCTGAAATGTCTCAAGTGTGATGAGGTGAGCTGGCCAAACA 3720  
Qy 3159 AAGCTCAGTCCCTGCGCACAGCCCGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3218  
Dh 3721 ACTAAAGCCCATCTGCATCAACC-----CACAGGTAAAGCCAG 3759  
Qy 3219 GAGCTAAGGTTCCCGCGCGTTCCTTATCTTCAGGAGAAAGTGGCCGTAAAGAT 3278  
Dh 3760 GTGCTTAGGTGCCAGCTGTTAATTTTCACCCCATCAGAGAGAGGAGCAACAAAGAC 3819  
Qy 3279 CTCACACCTTTTGGCATAACTCTCCAAAGCTTTTAAAAAAAATCTTTTGTGTTGAA 3338  
Dh 3820 CTGTCATGTAATACAAAGTTTACGTCAAGAAAGTTTATAGAACAA-----CTGAA 3887  
Qy 3339 GAAGGAAGTTTGAAGCTCAGCAAGAAACAGGCTGAAGATACATATCAATGATCTTCG 3398  
Dh 3868 GAAATAA-----GTGGTAAGAACCATACAGAACATATTTCTGTGGGCTCATCT 3918  
Qy 3399 CAGCTTCT 3458  
Dh 3919 TTACATTTCTAGTCT 3978  
Qy 3459 CCAAGTGTACTGTGATTAATTTTTCAGATTTCTGTCACAGTAAATTTCTTTCACAGATTC 3518  
Dh 3979 CCATCAGCTAAATATGTACAACTTGTGACCTCCAGCATATGATGATTTCTTTCAGGGTCC 4038

```
QY 3519 AGATTCTTAGCAATTCCTCTTTTGACTGACGTCACATGTCAGTGCAGATGAGAGCCG 3578
      || || || || || || || || || || || || || || || || || || || || ||
Db 4039 AGCATTCGTAGCAATTCCTCTTTTGACTGACGTCACATGTCAGTGCAGATGAGAGCCG 4098
QY 3579 CAGAGGATTCCTGTCAGCATGTCGAAACCAACCTAGGATGGGAGATGAGAGCCG 3638
      || || || || || || || || || || || || || || || || || || || || ||
Db 4099 TCCTCTCAGAGCCCTGGGAGAGTCCCTGATTCACATGGGAGATGAGAGAGAGCT 4158
QY 3639 ACCATTTGAACTGATCAGTATAGCTTGGGGTCCATGACCAATGTCGAGGGCCGA 3698
      | || | || | || | || | || | || | || | || | || | || | || |
Db 4159 TCAGGAGATGAGAGATCATAGTCAACATGGCCCTGGGAGAGCTTTGAAGCATCTCTA 4218
QY 3699 GCGTTATATGCTACAGCTAGTATATTTCTCTCCAGCAGAGAACTTCCAGAGAT 3758
      || || || || || || || || || || || || || || || || || || || ||
Db 4219 ATCAAGCTTTTACTGTCTCA-----TCCTCTGTGAGCATGAGAGATTTCTCAAGAG 4272
QY 3759 CAGGGGATCGCGGCTCCTGATGCTGCTGACAGTGGCCCTGGGAGCTGGAGCTCATGC 3818
      || || || || || || || || || || || || || || || || || || || ||
Db 4273 CATATCATTT-----ATGAAAGCAGCTGACAGTGGTCTGTGAGATTTGAGCTTGGTGT 4323
QY 3819 TCAAGTGGCTCCCATGATATATATACAGAGATCCAGCAGAGAAAGCTGGGA 3871
      || || || || || || || || || || || || || || || || || || || ||
Db 4324 TCAAGCAGCTCCCATGACATCTCCAAAGCCTTCCAAACCAAAAGAGCTGGGA 4376

RESULT 4
US-10-046-935-1879/c
; Sequence 1879, Application US/10046935
; Patent No. US2002015601A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1879
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-046-935-1879

Query Match          6.1%; Score 399; DB 9; Length 399;
Best Local Similarity 100.0%; Pred. No. 1.8e-81;
Matches 399; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Db 99 CTCTACCAAGATTGACCAAAATGTTATCTAGCAATTAAATGTAGAAACATAGAGCTG 40
      || || || || || || || || || || || || || || || || || || || || ||
QY 6321 CTAACTCAGTTCCCTCTGTGATGATCAAGTGCAGATGT 6359
      || || || || || || || || || || || || || || || || || || || || ||
Db 39 CTAATCTCAGTTCCCTCTGTGATGATCAAGTGCAGAAATGT 1

RESULT 5
US-09-878-178-1879/c
; Sequence 1879, Application US/09878178
; Patent No. US2002017552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1879
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-878-178-1879

Query Match          6.1%; Score 399; DB 9; Length 399;
Best Local Similarity 100.0%; Pred. No. 1.8e-81;
Matches 399; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 5961 ACTCAATGTGAGAGACTGCTGCTTGGATTTTTTTCATTAAATTCAGCTGATCATTA 6020
      || || || || || || || || || || || || || || || || || || || || ||
Db 399 ACTCAATGTGAGAGACTGCTGCTTGGATTTTTTTCATTAAATTCAGCTGATCATTA 340
QY 6021 TTGATCAGTATGATTAACGTAATAGCTTCAAAATTTTAAAGTGAATTCAGTGTTTT 6080
      || || || || || || || || || || || || || || || || || || || || ||
Db 339 TTGATCAGTATGATTAACGTAATAGCTTCAAAATTTTAAAGTGAATTCAGTGTTTT 280
QY 6081 CACTGTATCAACAATGTCACTGCTTTATTTAATATTCCTCTGTATATATGAGCAATTTG 6140
      || || || || || || || || || || || || || || || || || || || || ||
Db 279 CACTGTATCAACAATGTCACTGCTTTATTTAATATTCCTCTGTATATATGAGCAATTTG 220
QY 6141 TCTACTTGTCTTATTAATAGCTGATTTGATTTGATTTTACATTAATATGATTAAT 6200
      || || || || || || || || || || || || || || || || || || || || ||
Db 219 TCTACTTGTCTTATTAATAGCTGATTTGATTTGATTTTACATTAATATGATTAAT 160
QY 6201 TTGCAGTTTATTAATAGCTGATTTGATTTGATTTTACATTAATATGATTAAT 6260
      || || || || || || || || || || || || || || || || || || || || ||
Db 159 TTGCAGTTTATTAATAGCTGATTTGATTTGATTTTACATTAATATGATTAAT 100
QY 159 TTGCAGTTTATTAATAGCTGATTTGATTTGATTTTACATTAATATGATTAAT 100
QY 6261 CTCTACCAAGATTGACCAAAATGTTATCTAGCAATTAAATGTAGAAACATAGAGCTG 6320
      || || || || || || || || || || || || || || || || || || || || ||
```

```
US-09-911-826a-3
; Sequence 3, Application US/09911826A
; Patent No. US2002014316A1
; GENERAL INFORMATION:
; APPLICANT: Rolin, Daniela and Pham, Nam
; TITLE OF INVENTION: RAS Activator Nucleic Acid Molecules, Polypeptides and
; FILE REFERENCE: DDW-5001-US
; CURRENT APPLICATION NUMBER: US/09/911,826A
; CURRENT FILING DATE: 2002-02-26
```

```

; PRIOR APPLICATION NUMBER: PCT/CA00/00042
; PRIOR FILING DATE: 2000-01-20
; PRIOR APPLICATION NUMBER: 2,259,830
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 799
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-911-826a-3
```

```

Query Match      6.0%; Score 394.8; DB 10; Length 799;
Best local Similarity 74.2%; Pred. No. 2.3e-80;
Matches 582; Conservative 0; Mismatches 187; Indels 15; Gaps 6;
```

```

QY 4034 AACGTGGGCGCTCTTCACAGCTTACTGAGGAGAGAGACTGAGAAAGTGACACAGGCAACAT 4093
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 15  AAGCTGGAGCTCCACCGCGGTGGCGGCGCTCTAGAACTAGTGGATGCCCGGGCTGACAG 74
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 4094 AAAGCGGAGGGGTGGAAAGAGATGTTTCATTGAAGCCGAAAGCAAGTACCTTAACGTCTGT 4153
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 75  GAATTCAGGCGGTGGAGAGATGTCTCCGCTGAGGCAAGAGACAGAGCAATGCTGCTCCGT 134
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 4154 GACTACGGAAGAAACCAAGCCTGTCCCATGCTGCCACATAGCTGTGCTGATCAAGTAC 4213
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 135  GATACAGAGAAAGCAACCTGTCTCTATGCTGCTGCCACATAGCTGTGACGCCGAGAC 184
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 4214 TACAAAGGGGCTATTGACAGAAAGAGAGGAGATGAGAGCCCCCGCCACCCCTCC 4273
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 195  TACCAAGGGGACTATGCTGACAGAAAGAGAGATGACCGGAGCCGCTCCACACCTCC 254
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 4274 CGGCTACATTTGSAATTTCTTCTACTGCTTTCAGAAAGGCGACTCCCATCCACAGAGAA 4333
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 255  AGGCTACGTGGGCAATCCCATGTGCGGCTTCCAGAAAGGCGCTTCCACCGGCGAGGAA 314
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 4334 ACCGCGGAGTACACGCTGCGCTTCAGAGATGCGGATGCTGCGCATCTCCGACAC 4393
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 315  GCCCGCGGATTAACAGCTGGCGCTGACGAGGCTCCCGCATGTGGAGCGGCCACTGAGGC 374
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 4394 AGCTGGGCTTTCATCCGTACAGACAGCCACATGGGCAATCCACAGCAGCAGGCGCTGTAA 4453
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 375  CCGGCGCAC-----GGGCGACAGCGCGCTGACAGCGCAGCAGCGCGGCGAG 425
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 4454 CAACACCTCAGTGGCATTAACCGAAGCGTGTGACCCCGCGCTGCGCCCTTATCAGTCCCA 4513
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 426  CAAGCCACAGTGGACAGAGCCCGACGACGACCCCGCTGCGCGCTTCCAGCGCAGG 485
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 4514 AGGGTTTTCACCGAGAGAGATGAGTGAACAAGTTTCTGCTGTTTGGAGCAGACACTT 4573
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 486  CTTCGACGAGAGCGAGAGAGAGATGAACAAGTGTCTGCTGTTTGGAGCGCAGGC-- 543
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 4574 TTCTGGAAGACAGAGCGACCACTGAAAGAGAGACACAGAAGACGTCTGAGCATTTGA 4633
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 544  TCGTTGATCCACAGTGAAGCCACC-CAAAAGAGAGACAAAGAAGACGTCACAGCTTGA 602
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 4634 GCCTGGAATCACAATCTGAGAGAGCGTGAACAGATTGGCTCTGCTCCCTGCTTAAAG 4693
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 603  GCCCTGGACAGACACA-TCTAAGAGTGTGACACAGATTGGCTCTGCTCCCTTAAAG 660
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 4694 CAGCATGGGCTTCTCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4753
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 661  CAGATGGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 4754 AATTGCGCTGGCACTTTTCAGACTTTGTTGCTTGAATAAGCAGAGTGCAGAACTCTGAG 4813
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 721  AATTGCGCTGGCACTTTTCAGAGAC-TTGTTGCTTGAATAAGCAGAGCCAGAGCCCTGAG 779
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 4814 CTCG 4817
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 780  CTCG 783
```

```

RESULT 7
US-10-046-935-1196/c
; Sequence 1196, Application US/10046935
; Patent No. US20020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secret, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527c1
; CURRENT APPLICATION NUMBER: US/10/046,935
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1196
; LENGTH: 400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-046-935-1196

Query Match      5.9%; Score 388; DB 9; Length 400;
Best local Similarity 99.8%; Pred. No. 5.8e-79;
Matches 399; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 5961 ACTCAATGTGAGAACCTGCTGTGATTTTTCATTAATTCAGCTGATCATA 6020
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 400  ACTCAATGTGAGAACCTGCTGTGATTTTTCATTAATTCAGCTGATCATA 341
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 6021 TTGATCAGTGAATACGTAATAGCTTCAATTTTAAAGTGAATGAGAG-TTTT 6079
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 340  TTGATCAGTGAATACGTAATAGCTTCAATTTTAAAGTGAATGAGAGTGTGTTT 281
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 6080 TCACTGATCAACAATGTCAGTCTTTTATTAATTTCTTCTGTATCATGCGACTT 6139
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 280  TCACTGATCAACAATGTCAGTCTTTTATTAATTTCTTCTGTATCATGCGACTT 221
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 6140 GTCTACTGCTTATTCATTTGCAATTTGCAATTTGTAATTTTACATGTAATGCAATTA 6199
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 220  GTCTACTGCTTATTCATTTGCAATTTGCAATTTGTAATTTTACATGTAATGCAATTA 161
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 6200 TTGGCGAGTTTATTAATGCTATGACCTCATGTGCTATTAAGAAAGACAGAAATCTA 6259
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 160  TTGGCGAGTTTATTAATGCTATGACCTCATGTGCTATTAAGAAAGACAGAAATCTA 101
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 6260 GCTCTACCAACAATGTCAGCAATGTTATCTAAGCATTAAGTAAATGAGAATAGACT 6319
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 100  GCTCTACCAACAATGTCAGCAATGTTATCTAAGCATTAAGTAAATGAGAATAGACT 41
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 6320 GCTAATCTAGTTGCTGTGTGATGTCAGTGAAGTGT 6359
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 40  GCTAATCTAGTTGCTGTGTGATGTCAGTGAAGTGT 1

RESULT 8
US-09-878-178-1196/c
; Sequence 1196, Application US/09878178
; Patent No. US2002017752A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secret, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1196
; LENGTH: 400
```



```
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-878-178-1196

Query Match
Best Local Similarity 99.8%; Score 388; DB 9; Length 400;
Matches 399; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 5961 ACTCAATGTGAGAACTGCTGCTGATTTTTCATTAATTCAGTGCATCA 6020
    |||||||
DB 400 ACTCAATGTGAGAACTGCTGCTGATTTTTCATTAATTCAGTGCATCA 341

OY 6021 TTGATCAGTAGATAACGTAAATAGCTTCAATTTTAAAGTGCAATGCA 6079
    |||||||
DB 340 TTGATCAGTAGATAACGTAAATAGCTTCAATTTTAAAGTGCAATGCA 281

OY 6080 TCACGTATCAACAAGTCAAGTCTTATTTATTAATTCCTTCATCAGGCA 6139
    |||||||
DB 280 TCACGTATCAACAAGTCAAGTCTTATTTATTAATTCCTTCATCAGGCA 221

OY 6140 GTCCTCTGCTTATTTACATGTCAATTTGTAATTTTACATGTAATGCA 6199
    |||||||
DB 220 GTCCTCTGCTTATTTACATGTCAATTTGTAATTTTACATGTAATGCA 161

OY 6200 TTTGCCAGTTTATTTATAGGCTATGAGCCTCATGTGCATATGAAAGACA 6259
    |||||||
DB 160 TTTGCCAGTTTATTTATAGGCTATGAGCCTCATGTGCATATGAAAGACA 101

OY 6260 GCTCTACCAAGTTCGCAAAATGTTATCTAAGCATTAAGTAATGTGACAT 6319
    |||||||
DB 100 GCTCTACCAAGTTCGCAAAATGTTATCTAAGCATTAAGTAATGTGACAT 41

OY 6320 GCTAATCTCAGTTCGCTCTGTGATGCTCAAGTGACAGAAATGT 6359
    |||||||
DB 40 GCTAATCTCAGTTCGCTCTGTGATGCTCAAGTGACAGAAATGT 1

RESULT 9
US-09-733-607-15/C
; Sequence 15, Application US/09733607
; Patent No. US20020042054A1
; GENERAL INFORMATION:
; APPLICANT: Tononi, Giulio
; APPLICANT: Cirelli, Chiara
; APPLICANT: Shaw, Paul J.
; APPLICANT: Greenspan, Ralph J.
; TITLE OF INVENTION: Vigilance Nucleic Acids and Related
; FILE REFERENCE: P-NI 4447
; CURRENT APPLICATION NUMBER: US/09/733,607
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US 09/456,785
; PRIOR FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 263
; TYPE: DNA
; ORGANISM: raltus
US-09-733-607-15

Query Match
Best Local Similarity 91.6%; Score 227.8; DB 10; Length 263;
Matches 241; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

OY 526 TGGTGAAGAAGCAGGACCATAGTGTAAATAGTGTGAGAGCTGGACTCCTG 585
    |||||||
DB 263 TGGTGAAGAAGCAGGACCATAGTGTAAATAGTGTGAGAGAGCTGGACTC 204

OY 586 TGATTTCATGATCTGTGAGAGTCAATTCAGATGGAAGCAAAATCTGTGCA 645
    |||||||
DB 203 TGATTTCATGATCTGTGAGAGTCAATTCAGATGGAAGCAAAATCTGTGCA 144
```

```
OY 646 TGGCAATAGTTTGTCTCTCTCTACATGACCAAGAATACATGAAGAGATGATGA 705
    |||||||
DB 143 TGGCAACAGCTTTGGTGTCTTGGCCACACATGACACAAAGATACATGAAGATGATGA 84

OY 706 GACCAAGGTGATGACTGCCAGTTTGTCTGATAGCCACAGCAAGATTACTGCCATTC 765
    |||||||
DB 83 GACCAAGGTGATGACTGCCAGTTTGTCTGATAGCCACAGCAAGATTACTGCCATTC 24

OY 766 TCATATCAAGTAGAAAAAACAATG 788
    |||||||
DB 23 TCATATCAAGTAGAAAAAACAATG 1

RESULT 10
US-09-864-761-27564/C
; Sequence 27564, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecmlica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 27564
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004622.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.65
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.6
```



```

1 CURRENT APPLICATION NUMBER: US/09/864,761
2 CURRENT FILING DATE: 2001-05-23
3 PRIOR APPLICATION NUMBER: US 60/180,312
4 PRIOR FILING DATE: 2000-02-04
5 PRIOR APPLICATION NUMBER: US 60/207,456
6 PRIOR FILING DATE: 2000-05-26
7 PRIOR APPLICATION NUMBER: US 09/632,366
8 PRIOR FILING DATE: 2000-08-03
9 PRIOR APPLICATION NUMBER: GB 24263.6
10 PRIOR FILING DATE: 2000-10-04
11 PRIOR APPLICATION NUMBER: US 60/236,359
12 PRIOR FILING DATE: 2000-09-27
13 PRIOR APPLICATION NUMBER: PCT/US01/00666
14 PRIOR FILING DATE: 2001-01-30
15 PRIOR APPLICATION NUMBER: PCT/US01/00667
16 PRIOR FILING DATE: 2001-01-30
17 PRIOR APPLICATION NUMBER: PCT/US01/00664
18 PRIOR FILING DATE: 2001-01-30
19 PRIOR APPLICATION NUMBER: PCT/US01/00669
20 PRIOR FILING DATE: 2001-01-30
21 PRIOR APPLICATION NUMBER: PCT/US01/00665
22 PRIOR FILING DATE: 2001-01-30
23 PRIOR APPLICATION NUMBER: PCT/US01/00668
24 PRIOR FILING DATE: 2001-01-30
25

```

```

: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Weosheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
: FILE REFERENCE: Aemica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312

```

PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1  
SEQ ID NO 5303  
LENGTH: 471  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC004227.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.8  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.7  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.3  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.7  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.2  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.5  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.6  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4  
US-09-864-761-5303

Query Match 1.6%; Score 104.6; DB 10; Length 471;  
Best Local Similarity 69.1%; Pred. No. 2,5e-14;  
Matches 143; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 1155 A A A A T G G T G G A C C T A A G C T G T T G A T A T C G C G T G C T G C T T A A G C A A A A G A A G A 1214  
DB 471 A G A T A A A T G C A T C T C G G T A T T G A T A T T G C T G T G C A A A G C T A A G T G A G A 412  
QY 1215 T T G A T A C G T T A C A A A C A A A C C G A A G C T T T C C T T T A T C T T A C T T G A G C 1274  
DB 411 C A G T G T G T G C A A A A G C T T C C C G A G T C C C C T C T A C A T T A G C C T T A A T G A G G 352  
QY 1275 T C T G A A A G A G A T T T G A A T C T T T G T T G A C A G T A G A T T A G A T T A G A C A A C A A C T G A A 1334  
DB 351 A G T A A A G A G A T T T G T A T T T T T T G T T G A A G A G T A G A C C T G T A G C A A A C T G C T G A T 292  
QY 1335 G C A G C T T G A A A C G G G G G A T C A G A T A 1361  
DB 291 T C A G A C T G A A A C G T G T G A T C A G T A 265

RESULT 13  
US-09-954-456-1970  
Sequence 1970, Application US/09954456  
Patent No. US20020115057A1  
GENERAL INFORMATION:  
APPLICANT: Young, Paul  
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C  
FILE REFERENCE: 689290-76  
CURRENT FILING DATE: 2001-09-18  
PRIOR APPLICATION NUMBER: US/60/233,617  
PRIOR FILING DATE: 2000-09-18  
PRIOR APPLICATION NUMBER: US/60/234,052  
PRIOR FILING DATE: 2000-09-20  
PRIOR APPLICATION NUMBER: US/60/234,923  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: US/60/235,134  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: US/60/235,637  
PRIOR FILING DATE: 2000-09-26  
PRIOR APPLICATION NUMBER: US/60/235,638  
PRIOR FILING DATE: 2000-09-26  
PRIOR APPLICATION NUMBER: US/60/235,711  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: US/60/235,720  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: US/60/235,840  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: US/60/235,863  
PRIOR FILING DATE: 2000-09-27  
NUMBER OF SEQ ID NOS: 2276  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1970  
LENGTH: 5900  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-954-456-1970

Query Match 1.6%; Score 104.6; DB 10; Length 5900;  
Best Local Similarity 55.2%; Pred. No. 9,3e-14;  
Matches 228; Conservative 0; Mismatches 179; Indels 6; Gaps 1;

QY 2355 A A C C G A A A C A T T T G G T A G A C A T C G A A A T T C T G A G A A A C A A A C C A G C T A A G A G G 2414  
DB 1235 A A T A G A G T C A G C A T T T G G G G C A G A G A T T C T G C T G C A G C C A G C T G G G C A A G C G A 1294  
QY 2415 A T G A A T C A T T A A G C A T T T C A T C A G A T A G C A C T G C A C T G T A G G A A T G C A A G A T T T 2474  
DB 1295 G T C A G C T G T G A A A A A T T C A A A A T T G C G G T C A C T G C A A A G C C C A G A G A A A C T G 1354  
QY 2475 A A C T C A A T G T T G C A A T A T C A G A T C A G T G C C T A A C C T G C A C C A G C A G A C T G C G A A G 2534  
DB 1335 A A T T C T T T T T G C A T T G A T G A T G G T C T C A C A C T G C T T C T G A G T G A G T G C A G 1414  
QY 2535 A C C T G G A A A A C T T C C A A T A A T A C G A A A A C T A T T T A A G A T C C A A G C T G T T 2594  
DB 1415 A C C T G G A A A A T C C C T G G A A G T T A A G A A C T T T C T G A C A C T G A A A G T T T A C A 1474  
QY 2595 G A T C C T T C A A A A C A T G C C A A A T A T C T A A T G T C T A A T G T C A A A T G T C A A A T T A C A C T 2654  
DB 1475 G A T C C T T C C T A A T T C A A A A G C C T A C A G A G A T C A T T C A - - - - - A A A A G A T G A A G C C A 1528  
QY 2655 C C C A A T A A C C C T A T T C C C A G T A T C A A A A A G A G A T C A C C T T C C T T C A G A A G A A A T 2714  
DB 1529 C C A A A A T C C C T T C A T G C C C T T A T T C C T T A A A G A T A C A T T A T T A T T A T T A G A A G A A T 1588  
QY 2715 G A C T C A A A A G T A G A C G G G C T G C A A T T T T G A G A G A C T A A G A T G A T G C A A A 2767  
DB 1589 A A A A C T T T T T G A A A T C T T G A A T T T T G A A A A G C T G A T A T A T G A T G C A G A 1641





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 10:34:29 : Search time 203 Seconds

(without alignments)  
922.428 Million cell updates/sec

Title: US-09-911-826a-1

Perfect score: 6568  
Sequence: 1 ctgcgcacgagagagattg.....aaaataataataacat 6568

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/PTCUTS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	62.4	1.0	7218	1 US-08-232-463-14	Sequence 14, Appl
C 2	60.8	0.9	5398	3 US-09-356-952-11	Sequence 11, Appl
C 3	58.2	0.9	43676	3 US-09-356-952-12	Sequence 12, Appl
C 4	51.6	0.8	7218	1 US-08-232-463-14	Sequence 14, Appl
C 5	51.6	0.8	19124	2 US-08-487-826a-13	Sequence 13, Appl
C 6	48	0.7	3275	4 US-09-370-838-151	Sequence 151, Appl
C 7	47	0.7	1474	4 US-08-821-994-64	Sequence 64, Appl
C 8	47	0.7	6200	4 US-09-439-923-1	Sequence 1, Appl
C 9	45.8	0.7	2550	6 5258287-23	Patent No. 5258287
C 10	45	0.7	2805	4 US-09-653-839-9	Sequence 9, Appl
C 11	44.8	0.7	1927	4 US-09-336-536-66	Sequence 66, Appl
C 12	44.6	0.7	140	1 US-08-628-417-5	Sequence 5, Appl
C 13	44.6	0.7	240	1 US-08-628-417-6	Sequence 6, Appl
C 14	44.4	0.7	1582	3 US-08-545-196a-10	Sequence 10, Appl
C 15	44.4	0.7	1582	3 US-08-545-196a-12	Sequence 12, Appl
C 16	44.4	0.7	2246	4 US-09-363-708-3	Sequence 3, Appl
C 17	44.2	0.7	1046	1 US-08-361-467b-4	Sequence 4, Appl
C 18	44.2	0.7	1046	1 US-08-484-332c-4	Sequence 4, Appl
C 19	44.2	0.7	3268	3 US-09-356-952-13	Sequence 13, Appl
C 20	43.8	0.7	1508	4 US-09-039-046-1	Sequence 1, Appl
C 21	43.8	0.7	1279	3 US-09-248-335-25	Sequence 25, Appl
C 22	43.8	0.7	2665	4 US-08-971-089-5	Sequence 5, Appl
C 23	43.8	0.7	3410	4 US-09-020-956-110	Sequence 110, Appl
C 24	43.8	0.7	3410	4 US-09-030-607-110	Sequence 110, Appl
C 25	43.8	0.7	3410	4 US-09-605-785-110	Sequence 110, Appl
C 26	43.8	0.7	3410	4 US-09-439-313-110	Sequence 110, Appl
C 27	43.8	0.7	3410	4 US-09-352-616a-110	Sequence 110, Appl

C 28	43.8	0.7	3410	4 US-09-602-877a-100	Sequence 100, Appl
C 29	43.8	0.7	3410	4 US-09-232-149a-110	Sequence 110, Appl
C 30	43.6	0.7	1027	4 US-09-465-558-57	Sequence 57, Appl
C 31	43.6	0.7	1098	3 US-09-248-335-35	Sequence 35, Appl
C 32	43.6	0.7	1872	4 US-09-801-052-1	Sequence 1, Appl
C 33	43.4	0.7	144	1 US-08-702-344-26	Sequence 26, Appl
C 34	42.8	0.7	1100	4 US-07-861-458c-4	Sequence 4, Appl
C 35	42.8	0.7	1117	4 US-09-247-373b-33	Sequence 33, Appl
C 36	42.6	0.6	1069	4 US-09-372-422a-7	Sequence 7, Appl
C 37	42.6	0.6	1319	2 US-08-504-459-7	Sequence 24, Appl
C 38	42.6	0.6	1493	1 US-08-820-820-24	Sequence 24, Appl
C 39	42.6	0.6	1493	1 US-08-593-535-24	Sequence 24, Appl
C 40	42.6	0.6	6124	4 US-08-213-419b-3	Sequence 3, Appl
C 41	42.4	0.6	1882	4 US-09-370-253-1	Sequence 1, Appl
C 42	42.2	0.6	2671	6 5168051-9	Patent No. 5168051
C 43	42	0.6	1114	4 US-09-152-060-41	Sequence 41, Appl
C 44	42	0.6	1359	4 US-09-387-574-11	Sequence 11, Appl
C 45	42	0.6	1359	4 US-09-668-096-11	Sequence 11, Appl

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BEYNT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZpct-F1s  
; US-08-232-463-14  
Query Match 1.0%; Score 62.4; DB 1; Length 7218;

RESULT 4  
US-08-232-463-14



```
Db 15737 TTTATTTTATATATTTTTTTCTTTTTTTTGTGTTTATGATATATATTTTTTTTTT 15678
QY 6268 ACAAGTTGCACAATGTATATCAGCATTAAGTAATGTAGACATAGAGCTCATCT 6327
Db 15677 TAATGTTTTTTTTTCTCTCTTTGTTTTTATTTTATTTTATTCATTTTTTTTTATAT 15618
QY 6328 CAGTTCGCTCGTGCATGTCAGATGACAGATGTACATTAATGATGATTCCTCATCTT 6387
Db 15617 AAAATTTTTTTTTTATTTTTTTTGTGATATCTTT--TTCAATTTTATCTATCAAAAT 15560
QY 6388 TTGATCTACTGTACTGTATGCTTTTGAAGAAGATGCGAGCTGATCCCTT 6447
Db 15559 TATATTTATATATATTTTATATATTTTAAATAATTTTCTCTTTTTTTT--TTT 15503
QY 6448 TGTATTTTATACATATATGTACATATGTTATTTTGTGGAAGATGTAGAAAT 6507
Db 15502 TTTTATTTTATTAATTAATTTTTTTTTTATTTTCAATTTCTTTTTTTCATTTTATA 15443
QY 6508 GTACTATGTTTATGCTTCTACATCCAGTTGTACAGCTGGAATAATTAATATAC 6567
Db 15442 AATTTGTTTATATTTCTTTTAAATATAATACATATATATATAATATATATAAAA 15383

RESULT 6
US-09-370-838-151/c
; Sequence 151, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; EARLIER FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 151
; LENGTH: 3275
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-370-838-151

Query Match 0.7%; Score 48; DB 4; Length 3275;
Best Local Similarity 50.0%; Pred. No. 0.0095;
Matches 120; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 5514 TTTTAAGTTAATTTCTTTGATTTTGTAAATTTAGAGCTAGCTTTTGTTTTGT 5573
Db 3200 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3141
QY 5574 TTTTGTTTTATTTTAGAACAATTTAATCTGATAGCATTTGAGTAAGACCTTG 5633
Db 3140 TTTTATTTTATTTGATGTAATTTTATTTTATTTAGGACACAAAGCATTTGACCTG 3081
QY 5634 GATGTTTGAGTAATGCCAGCTTTTAACTGCTTTTCAAGACAGCTCCCTTTATTTG 5693
Db 3080 GATTTAGGAGTACGTTAGTAACTGAACGTTAATACAGTTAAGATTAAGTGAAC 3021
QY 5694 AATTTGATTTAGGAATTAACAAGCTTTTAAACGTATTAAGATCAAAACCTGGTTAG 5753
Db 3020 AATATACATTTACAGCTTGCTAGTAGGAGGCTACATCATCAATTTATTAAGTGCACATTTAG 2961

RESULT 7
US-08-821-994-64/c
; Sequence 64, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
```

```
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; EARLIER FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
; LENGTH: 1474
; TYPE: DNA
; ORGANISM: Brassica napus
US-08-821-994-64

Query Match 0.7%; Score 47; DB 4; Length 1474;
Best Local Similarity 71.3%; Pred. No. 0.011;
Matches 62; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 5514 TTTTAAGTTAATTTCTTTGATTTTGTAAATTTAGAGCTAGCTTTTGTTTTGT 5573
Db 1449 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1390
QY 5574 TTTTGTTTTATTTTAGAACAATTT 5600
Db 1389 TTTTATTTTATTTTAGAACAATATAT 1363

RESULT 8
US-09-439-923-1/c
; Sequence 1, Application US/09439923
; Patent No. 6426208
; GENERAL INFORMATION:
; APPLICANT: Emily D. Kakkis
; APPLICANT: Becky Tanamachi
; TITLE OF INVENTION: Recombinant Alpha-L-Iduronidase, Methods
; TITLE OF INVENTION: for producing and purifying the same and methods for
; FILE REFERENCE: 08000051US00
; CURRENT APPLICATION NUMBER: US/09/439,923
; EARLIER FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 6200
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1558)...(3516)
US-09-439-923-1

Query Match 0.7%; Score 47; DB 4; Length 6200;
Best Local Similarity 62.2%; Pred. No. 0.026;
Matches 74; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 5517 TTAAGTTATTTCTTTGATTTTGTAAATTTAGAGCTAGCTTTTGTTTTGT 5576
Db 3702 TTTCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3643
QY 5577 TGTATTTTATTAAGAAACAATTTAATCTGATAGCATTTGACAGTAAGACAGCTGG 5635
Db 3642 TTTTATTTTATTAATAATAATAATATATTTGCAAGGGGATGATGAGGAGGAGCATTTAG 3584
```

```
RESULT 9
5258287-23/c
; Patent No. 5258287
; APPLICANT: BAXTER, ROBERT C.; WOOD, WILLIAM I.
```





5574 TTTTGT TTTT TTAAGAGAAACATTTATAA 36

;; CITY: FALLS  
; STATE: VA

RESULT 15  
US-08-545-196B-12/C  
; Sequence 12, Application US/08545196B  
; Patent No. 6080577  
; GENERAL INFORMATION:  
; APPLICANT: MELKI, JUDITH  
; APPLICANT: MUMMICH, ARNOLD  
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE  
; TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP  
; STREET: PO BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA

[illegible]

Search completed: February 18, 2003, 17:42:37  
Job time : 556 secs

**BLANK PAGE**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 08:40:04 : Search time 1143 Seconds  
(without alignments)  
12940.617 Million cell updates/sec

Title: US-09-911-826a-1

Perfect score: 6568

Sequence: 1 ctgcgcacgtgagagattg.....aaataaataataacat 6568

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*

2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*

3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*

4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*

5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*

6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*

7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*

8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*

9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*

10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*

11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*

12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*

13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*

14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*

15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*

16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*

17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*

18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*

19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*

20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*

21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*

22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*

23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*

24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6568	100.0	6568	21	AAAS9383
2	6568	100.0	6568	24	ABN95037
3	2840.2	43.2	2890	20	AAV84550
4	2840.2	43.2	2890	22	ABA83333
5	1587	24.2	6383	24	ABK94949
6	1582.4	24.1	6403	21	AACT6867
7	471	7.2	481	22	AAI20515
8	459	7.0	467	22	AAI1617
9	404.8	6.2	4298	24	ABK34853

C	10	399.2	6.1	402	24	ABL8290	Human colon tumour
C	11	389.2	5.9	801	21	AAA59384	DNA clone encoding
C	12	388.2	5.9	403	24	ABL37607	Human colon tumour
C	13	289	4.4	465	23	ABL15285	Drosophila melanog
C	14	227.8	4.4	7754	23	ABL15284	Drosophila melanog
C	15	227.8	3.5	263	22	ABD07308	Rat vigilance nuclei
C	16	201.6	3.1	360	22	AAK18699	Human brain expres
C	17	201.6	3.1	360	22	AAK44631	Human bone marrow
C	18	201.6	3.1	360	22	AAI50611	Human bone marrow
C	19	201.6	3.1	360	24	ABSI8868	Probe #19297 used
C	20	191.4	2.9	858	21	AAC77279	Human genome-deriv
C	21	166.6	2.5	465	22	AAK05914	Human brain expres
C	22	166.6	2.5	465	22	AAK31552	Human bone marrow
C	23	166.6	2.5	465	22	AAI37431	Probe #6117 used t
C	24	166.6	2.5	465	24	ABSO6306	Human genome-deriv
C	25	152.2	2.3	158	22	AAE67529	Novel human polynu
C	26	145.2	2.2	323	21	AAA44665	Human secreted exp
C	27	104.6	1.6	471	22	ABA57274	Human foetal liver
C	28	104.6	1.6	471	22	ABA26837	Probe #5303 for ge
C	29	104.6	1.6	471	22	AAK05301	Human brain expres
C	30	104.6	1.6	471	22	AAK30893	Human bone marrow
C	31	104.6	1.6	471	22	AAI15451	Probe #5384 for ge
C	32	104.6	1.6	471	22	AAI36806	Probe #5492 used t
C	33	104.6	1.6	471	24	ABSO5641	Human genome-deriv
C	34	104.6	1.6	5900	24	ABL66660	Lung cancer relate
C	35	101	1.5	3373	21	AAAD00314	Rat Ras signalling
C	36	99.2	1.5	3013	21	AAAD00318	Human Ras signalling
C	37	96.4	1.5	504	22	AAE93632	CDNA encoding SRP
C	38	95.8	1.5	1897	22	AAK51839	Human polynucleoti
C	39	95.8	1.5	1897	22	AAI58217	Human polynucleoti
C	40	95.8	1.5	2536	24	ABK37338	CDNA sequence #129
C	41	93.8	1.4	1666	21	AAAD00317	Rat Ras signalling
C	42	89.2	1.4	3384	21	AAAD00315	Human Ras signalling
C	43	87.8	1.3	182	22	ABK69505	Human foetal liver
C	44	87.8	1.3	182	22	ABA36758	Probe #15224 for g
C	45	87.8	1.3	182	22	AAK18106	Human brain expres

#### ALIGNMENTS

RESULT 1	AAAS9383	standard; DNA: 6568 BP.
ID	AAAS9383	
AC	AAAS9383;	
XX		
DT	07-NOV-2000	(first entry)
XX		
DE	DNA encoding a murine guanine nucleotide releasing factor 4.	
XX		
KW	guanine nucleotide releasing factor 4; GRF-4; Ras activator;	
KW	Nedd4 ubiquitination; cell metabolism; cell proliferation; cancer;	
KW	cell differentiation; cell transformation; neuronal disorder; ss.	
XX		
OS	Mus sp.	
XX		
FH		
FT	Key	Location/Qualifiers
FT	CDS	63..4562
FT		/tag- a
FT		/product- "guanine nucleotide releasing factor 4"
XX		
PN	WO200043510-A2.	
XX		
PD	27-JUL-2000.	
XX		
PF	20-JAN-2000; 2000MO-CA00042.	
XX		
PR	20-JAN-1999; 99CA-2259830.	
XX		
PA	(HSCR-) HSC RES & DEV LP.	
XX		
PI	Rotin D, Pham N;	

XX WPI; 2000-499228/44.  
DR P-PSDB; AAB07792.  
XX  
PT Nucleic acids encoding guanine nucleotide releasing factor-4 useful for  
the treatment of cancers and neuronal disorders -  
XX  
PS Claim 2; Fig 19A; 89pp; English.  
XX  
CC The present sequence encodes a murine guanine nucleotide releasing  
CC factor (GRF)-4 (Ras activator) polypeptide. GRF4 activates Ras both  
CC in vitro and in vivo. It directly binds cyclic adenosine monophosphate  
CC (cAMP) directly via its cAMP-BD (cAMP/guanine monophosphate (cGMP)  
CC binding domain). GRF4 directly connects cAMP-generating (e.g. G protein  
CC coupled receptors) or cGMP-generating pathways to Ras. GRF4 activates  
CC Ras in response to elevation of intracellular cAMP and/or cGMP. GRF4  
CC is a target for Nedd4 ubiquitination as it binds Nedd4. Activation of  
CC the Ras signalling pathway controls numerous cellular functions, such as  
CC cell metabolism, proliferation, differentiation and transformation.  
CC Therefore modulation of Ras activity may provide a mechanism for  
CC controlling diseases. GRF4 polynucleotides and polypeptides may be used  
CC in the treatment of diseases associated with inappropriate GRF4  
CC expression and activity such as cancers and neuronal disorders. The  
CC GRF4 polypeptides may be used as antigens in the production of  
CC antibodies against GRF4 and in assays to identify modulators (agonists  
CC and antagonists) of GRF4 expression and activity. The anti-GRF4  
CC antibodies and GRF4 antagonists may also be used to down regulate GRF4  
CC expression and activity. Inhibition of Ras can reduce cellulose  
CC proliferation and cancers.  
XX  
SQ Sequence 6568 BP; 1974 A; 1400 C; 1463 G; 1731 T; 0 other;  
Query Match 100.0%; Score 6568; DB 21; Length 6568;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CTTGGCATCGAGAGATTGGTACATGATGTGTAATTCAGTTCAGCATATGTTCTTCA 60  
DB 1 CTTGGCATCGAGAGATTGGTACATGATGTGTAATTCAGTTCAGCATATGTTCTTCA 60  
QY 61 TTATGAAACCACTAGCAATCCAGCTAACATGAGATTATGGGCCAGCAGAAACACT 120  
DB 61 TTATGAAACCACTAGCAATCCAGCTAACATGAGATTATGGGCCAGCAGAAACACT 120  
QY 121 CACTTCCTCGAGATTTCACAAACCTGATCTAGACAGCTCCACCCAGAGGTGACC 180  
DB 121 CACTTCCTCGAGATTTCACAAACCTGATCTAGACAGCTCCACCCAGAGGTGACC 180  
QY 121 CACTTCCTCGAGATTTCACAAACCTGATCTAGACAGCTCCACCCAGAGGTGACC 180  
DB 121 CACTTCCTCGAGATTTCACAAACCTGATCTAGACAGCTCCACCCAGAGGTGACC 180  
QY 181 ACGTTCTTCTAGCCATTCAGAGTATAGTACTAGTATGATCTGGGACAGCACTTTT 240  
DB 181 ACGTTCTTCTAGCCATTCAGAGTATAGTACTAGTATGATCTGGGACAGCACTTTT 240  
QY 241 CTGATATCTACAGGCCACAGAAACGAGGCTGGTATATGAGTACTGAGTGGTCCAG 300  
DB 241 CTGATATCTACAGGCCACAGAAACGAGGCTGGTATATGAGTACTGAGTGGTCCAG 300  
QY 241 CTGATATCTACAGGCCACAGAAACGAGGCTGGTATATGAGTACTGAGTGGTCCAG 300  
DB 241 CTGATATCTACAGGCCACAGAAACGAGGCTGGTATATGAGTACTGAGTGGTCCAG 300  
QY 301 AAACAGCATGATTCGGAAGCAGCAGATGAGAAAGACACTTGAAGAGATAGATC 360  
DB 301 AAACAGCATGATTCGGAAGCAGCAGATGAGAAAGACACTTGAAGAGATAGATC 360  
QY 301 AAACAGCATGATTCGGAAGCAGCAGATGAGAAAGACACTTGAAGAGATAGATC 360  
DB 301 AAACAGCATGATTCGGAAGCAGCAGATGAGAAAGACACTTGAAGAGATAGATC 360  
QY 361 CTCTGATGACAGGACATTTGTAGAGATGCTAGAGAAAGACCCAAATGACGGACAG 420  
DB 361 CTCTGATGACAGGACATTTGTAGAGATGCTAGAGAAAGACCCAAATGACGGACAG 420  
QY 421 ATGATGACATTTGAACAACTCTTGAATTTATGACACAGTGGCTCTTTTGCCTATAGA 480  
DB 421 ATGATGACATTTGAACAACTCTTGAATTTATGACACAGTGGCTCTTTTGCCTATAGA 480  
QY 481 CAATGTCATGAGGAGAGAACTCTGCTGATGATGTTTCCAGTGGTGAAGAGAGAG 540  
DB 481 CAATGTCATGAGGAGAGAACTCTGCTGATGATGTTTCCAGTGGTGAAGAGAGAG 540  
QY 541 GGACCATAGTGTAAATGATGTTGAAGAGCTGCACTCTGTGATGATTTCAATGAT 600

DB 541 GGACCATAGTGTAAATGATGTTGAAGAGCTGCACTCTGTGATGATTTCAATGAT 600  
QY 601 CTGTGGAAGTGAATTCATTCAGATGGAAGAGAGAAATCTGCTGATGGGAATAGTTTG 660  
DB 601 CTGTGGAAGTGAATTCATTCAGATGGAAGAGAGAAATCTGCTGATGGGAATAGTTTG 660  
QY 661 GTGTCTCTCTCCATTCAGAGCAAAATATCATGAAGAGTGTGAAGAAAGGTGATG 720  
DB 661 GTGTCTCTCTCCATTCAGAGCAAAATATCATGAAGAGTGTGAAGAAAGGTGATG 720  
QY 721 ACTGCCAGTTTGTCTGATAGCCACAGCAAGATTACTGCCGATTTCTCAATCAAGTAGAA 780  
DB 721 ACTGCCAGTTTGTCTGATAGCCACAGCAAGATTACTGCCGATTTCTCAATCAAGTAGAA 780  
QY 781 AGAACAATGCAAAAAGTTGAAGAGAGAGAGAGATGTTTGTGAAGAAACCCAGAAC 840  
DB 781 AGAACAATGCAAAAAGTTGAAGAGAGAGAGATGTTTGTGAAGAAACCCAGAAC 840  
QY 841 TTGATGAACTGGAACAGAGAGAGGACACATTTGTCATCAAGGATCCTCAGAAAGTTAA 900  
DB 841 TTGATGAACTGGAACAGAGAGAGGACACATTTGTCATCAAGGATCCTCAGAAAGTTAA 900  
QY 901 CAATGCAATTTGTGGAAGAGCATTCAGTAGATCAATCAATTCATAGAGACTTCTGT 960  
DB 901 CAATGCAATTTGTGGAAGAGCATTCAGTAGATCAATCAATTCATAGAGACTTCTGT 960  
QY 961 TGACCTATAGAGACTTTCTTTCTAGCCCAATGAAATGGGCAAAAGTAAATGGAGTGT 1020  
DB 961 TGACCTATAGAGACTTTCTTTCTAGCCCAATGAAATGGGCAAAAGTAAATGGAGTGT 1020  
QY 1021 TTATGACCCGAGCTCAGAGGATAGGTTACACGGGTAGTATTTTGGGTAATATATC 1080  
DB 1021 TTATGACCCGAGCTCAGAGGATAGGTTACACGGGTAGTATTTTGGGTAATATATC 1080  
QY 1081 ACTTCAATGACTTTGAAGAGATCTGCAATGACTGATTTTGAAGAAATTTGAAGAA 1140  
DB 1081 ACTTCAATGACTTTGAAGAGATCTGCAATGACTGATTTTGAAGAAATTTGAAGAA 1140  
QY 1141 ATCTGGAAGAGAGAAATGGGTGACACCTAAGGCTGGTAAATTCGGGTGCTGCTA 1200  
DB 1141 ATCTGGAAGAGAGAAATGGGTGACACCTAAGGCTGGTAAATTCGGGTGCTGCTA 1200  
QY 1201 AAGCAAAAAGAGATTTGATGACCTTAAACAAACCATCCGAGAGACTCTTTGCCCTTTA 1260  
DB 1201 AAGCAAAAAGAGATTTGATGACCTTAAACAAACCATCCGAGAGACTCTTTGCCCTTTA 1260  
QY 1261 TCTTACTTGGAGGCTCTGAGAAGGATTTGGAAATCTTTGTGACAGTATGATTCAGGTA 1320  
DB 1261 TCTTACTTGGAGGCTCTGAGAAGGATTTGGAAATCTTTGTGACAGTATGATTCAGGTA 1320  
QY 1321 GCAAGCAACTGAAGCAGGCTTGAACGGGGGATCAGATATTAAGAAATGGCCAAA 1380  
DB 1321 GCAAGCAACTGAAGCAGGCTTGAACGGGGGATCAGATATTAAGAAATGGCCAAA 1380  
QY 1381 ACTTTGAAAACATTCACGTCTCAAAAGCTATGAAATCTTGAAGAAATACACATTTAT 1440  
DB 1381 ACTTTGAAAACATTCACGTCTCAAAAGCTATGAAATCTTGAAGAAATACACATTTAT 1440  
QY 1441 CTATCAGCTGAAAACCAATTTATTTATTTAAGAACTTCAACAGAAATGTGCAAG 1500  
DB 1441 CTATCAGCTGAAAACCAATTTATTTATTTAAGAACTTCAACAGAAATGTGCAAG 1500  
QY 1501 AGAAAAGAAATGTGCCCCCACCCTTCTAAATTTGTGACATTTAAAGGCAAGTGGCT 1560  
DB 1501 AGAAAAGAAATGTGCCCCCACCCTTCTAAATTTGTGACATTTAAAGGCAAGTGGCT 1560  
QY 1561 ACTCCATTCAGATCTTGTCTAGATGTAAGACAGGTGATAGGACTTGAAGAAAGTGAACA 1620  
DB 1561 ACTCCATTCAGATCTTGTCTAGATGTAAGACAGGTGATAGGACTTGAAGAAAGTGAACA 1620  
QY 1621 AAAAAAGTAAAGCCAAACACTGTGGAGAGAGCAAGCTGAAAGAAATGATCTGCAGAGA 1680

Dd 1621 AAAAAAGTAAACCCAAACAGTGTGGAGGAAAGAACAGCTGTAAGAAAAAGATATCTGCACAGA 1680  
Oy 1681 CTCGGATCAGTATCTTGGCCACAGAAACCATACATGATATTTGGATTTGGTCTAGTCTCAAG 1740  
Dd 1681 CTCGGATCAGTATCTTGGCCACAGAAACCATACATGATATTTGGATTTGGTCTAGTCTCAAG 1740  
Oy 1741 ATGACAGCATAGTAGGATTTAAGGACAGCAAAAGACATCCCAACTGCATTTGCCCTGTCAAGT 1800  
Dd 1741 ATGACAGCATAGTAGGATTTAAGGACAGCAAAAGACATCCCAACTGCATTTGCCCTGTCAAGT 1800  
Oy 1801 GAACCTTATCATCCCAAGTATCTCTGATTTATTCAGTCTACATCATGCTATTTAGACTTCA 1860  
Dd 1801 GAACCTTATCATCCCAAGTATCTCTGATTTATTCAGTCTACATCATGCTATTTAGACTTCA 1860  
Oy 1861 GTGCTACTCTCAGTCTGACAGATCAAGTCTTAAGGGTTTAAAGGCTGTATCGACAAAGCC 1920  
Dd 1861 GTGCTACTCTCAGTCTGACAGATCAAGTCTTAAGGGTTTAAAGGCTGTATCGACAAAGCC 1920  
Oy 1921 GCTACATCATGATCAGTAAAGACACTACAGCAAAAGAAAGTGTCTATTCAGGCTATCAGGG 1980  
Dd 1921 GCTACATCATGATCAGTAAAGACACTACAGCAAAAGAAAGTGTCTATTCAGGCTATCAGGG 1980  
Oy 1981 AGTTTGTCTTACTGCGCACCCCGGATCAATATTCTACTATGTAGAGTCTGTGTCAACCTG 2040  
Dd 1981 AGTTTGTCTTACTGCGCACCCCGGATCAATATTCTACTATGTAGAGTCTGTGTCAACCTG 2040  
Oy 2041 AGGAGTAACTCAACAAAGAGACATCCAGATCAGCTTCCAACTTGCAGACAGAAATC 2100  
Dd 2041 AGGAGTAACTCAACAAAGAGACATCCAGATCAGCTTCCAACTTGCAGACAGAAATC 2100  
Oy 2101 AACTGATGGAAGGATATTATCTGAAAAACACATGGAAGAAACACTCTTTGTTCAGATG 2160  
Dd 2101 AACTGATGGAAGGATATTATCTGAAAAACACATGGAAGAAACACTCTTTGTTCAGATG 2160  
Oy 2161 AAGATGCTCAGAGATTTGGAAGAGAGTCAATTTCCCTCTTCAAGCTCAGACACTGTGG 2220  
Dd 2161 AAGATGCTCAGAGATTTGGAAGAGAGTCAATTTCCCTCTTCAAGCTCAGACACTGTGG 2220  
Oy 2221 AAGTTGCAACAGAGCTCTCTATGCGAAATTTTGAACCTTTTGGCAACATTTGAACCTAG 2280  
Dd 2221 AAGTTGCAACAGAGCTCTCTATGCGAAATTTTGAACCTTTTGGCAACATTTGAACCTAG 2280  
Oy 2281 AATATATATGATGATTTTAACTCAGATCAAAAAACAGCTGTGGCAACCTGAAGAGAT 2340  
Dd 2281 AATATATATGATGATTTTAACTCAGATCAAAAAACAGCTGTGGCAACCTGAAGAGAT 2340  
Oy 2341 TTGGAAGAAGTCTTAAACCGGAAACATTTTGGGTAGCATCTGAATAATTTCAAGAAACAA 2400  
Dd 2341 TTGGAAGAAGTCTTAAACCGGAAACATTTTGGGTAGCATCTGAATAATTTCAAGAAACAA 2400  
Oy 2401 ACCAGCTGAAGAGATGAAGATCATTAAGCATTTTATCAAGATGACAGTGCACCTGTAGGG 2460  
Dd 2401 ACCAGCTGAAGAGATGAAGATCATTAAGCATTTTATCAAGATGACAGTGCACCTGTAGGG 2460  
Oy 2461 AATGCAAGAATTTTAACTCAATGTTTGAATCATCAGTGGCTTAAACCTGGACACAGTGG 2520  
Dd 2461 AATGCAAGAATTTTAACTCAATGTTTGAATCATCAGTGGCTTAAACCTGGACACAGTGG 2520  
Oy 2521 CAAGACTGGGAACGACTGGGAGAAACTTCCCAATTAATACGAAAAACTATTTCAAGATC 2580  
Dd 2521 CAAGACTGGGAACGACTGGGAGAAACTTCCCAATTAATACGAAAAACTATTTCAAGATC 2580  
Oy 2581 TCCAGAGACCTGTTTGAATCTTCAGAAACATGCAAAATATGTGTCAATAGTC 2640  
Dd 2581 TCCAGAGACCTGTTTGAATCTTCAGAAACATGCAAAATATGTGTCAATAGTC 2640  
Oy 2641 AAAATCTACACCTCCCATATTCCTTATTCAGATTATCAAAAAGATCTCACCTTCC 2700  
Dd 2641 AAAATCTACACCTCCCATATTCCTTATTCAGATTATCAAAAAGATCTCACCTTCC 2700  
Oy 2701 TTCACGAAGAATGACTCAAAAGTAGACGGGCTGTCAATTTTGAAGAGCTAAGGATGA 2760  
Dd 2701 TTCACGAAGAATGACTCAAAAGTAGACGGGCTGTCAATTTTGAAGAGCTAAGGATGA 2760

Oy 2761 TTGCAAAAAGAAATTCGTCACGTTGGCCGAATGCTTCAATGAACATGAGACCTGCCCTCA 2820  
Dd 2761 TTGCAAAAAGAAATTCGTCACGTTGGCCGAATGCTTCAATGAACATGAGACCTGCCCTCA 2820  
Oy 2821 TGTTCAGGACTCGGAAGAAGAAATGGCGGAGTTTGGGGTCTCTCAGCCAGGGTAGTACAA 2880  
Dd 2821 TGTTCAGGACTCGGAAGAAGAAATGGCGGAGTTTGGGGTCTCTCAGCCAGGGTAGTACAA 2880  
Oy 2881 ATGCAACATGCTAGATGTTTGTCTCAGACAGGTGTGTATAAAAAGCCGGGTACGTCGATG 2940  
Dd 2881 ATGCAACATGCTAGATGTTTGTCTCAGACAGGTGTGTATAAAAAGCCGGGTACGTCGATG 2940  
Oy 2941 CCTTCTCATATGCAAAAAAGCTTTATGAAGATGCCCAATTTGGCTCGAAAAGTGAACAG 3000  
Dd 2941 CCTTCTCATATGCAAAAAAGCTTTATGAAGATGCCCAATTTGGCTCGAAAAGTGAACAG 3000  
Oy 3001 ACCTTTCAATTTGAGCTAGAGTGAAGAAATGACGAGAGAGTCTTCAGACATTAATCTCTCAG 3060  
Dd 3001 ACCTTTCAATTTGAGCTAGAGTGAAGAAATGACGAGAGAGTCTTCAGACATTAATCTCTCAG 3060  
Oy 3061 GTGAGCCAGCAACCAACATTTGCTTAAGATCTGTGTGCAAAAAAGCCTGTCAATCCG 3120  
Dd 3061 GTGAGCCAGCAACCAACATTTGCTTAAGATCTGTGTGCAAAAAAGCCTGTCAATCCG 3120  
Oy 3121 AGACCTCTCAGTAGCTCCAGAGGGGACACACAGAAAGCTCAGTCCCTGCCACAG 3180  
Dd 3121 AGACCTCTCAGTAGCTCCAGAGGGGACACACAGAAAGCTCAGTCCCTGCCACAG 3180  
Oy 3181 CCCAGCAGGACCCACACAGCAGCATTAATCAACAGGAGACTACAGTGTCCCGCGGTG 3240  
Dd 3181 CCCAGCAGGACCCACACAGCAGCATTAATCAACAGGAGACTACAGTGTCCCGCGGTG 3240  
Oy 3241 CCTTTATCTTTCAGGAGAAAGATGCCGCTAAAGATCTCCACCTTTTGGCATAACT 3300  
Dd 3241 CCTTTATCTTTCAGGAGAAAGATGCCGCTAAAGATCTCCACCTTTTGGCATAACT 3300  
Oy 3301 CTCACACGCTTTAAAAAAATCTTTCTTGTCTGAAGAAAGATTTGGAAGCTCAC 3360  
Dd 3301 CTCACACGCTTTAAAAAAATCTTTCTTGTCTGAAGAAAGATTTGGAAGCTCAC 3360  
Oy 3361 AGAAACAGGCTGAAGATACAAATATCAATCTTCGAGCTTCTCTCTCTACTT 3420  
Dd 3361 AGAAACAGGCTGAAGATACAAATATCAATCTTCGAGCTTCTCTCTCTACTT 3420  
Oy 3421 CTCACAGAGTCTCCAAAGAAAGCTATATCTTGGCTCCAGTGTACTGTGGATTAAT 3480  
Dd 3421 CTCACAGAGTCTCCAAAGAAAGCTATATCTTGGCTCCAGTGTACTGTGGATTAAT 3480  
Oy 3481 TTTTCAATCTGGTCAAGAGTAAATTTCTTCAAGATCCAGTATTTGTAGCAATTCGCTT 3540  
Dd 3481 TTTTCAATCTGGTCAAGAGTAAATTTCTTCAAGATCCAGTATTTGTAGCAATTCGCTT 3540  
Oy 3541 TTGACTCAGTGGCAGTCTCACTGTCAGATGAGAGGGCCAGAGCAATTCGTGACGATCG 3600  
Dd 3541 TTGACTCAGTGGCAGTCTCACTGTCAGATGAGAGGGCCAGAGCAATTCGTGACGATCG 3600  
Oy 3601 TGGAAACAAACCTTAGGAGATGGGCAGATGAGAGGGGACCATGATTTGAACCTGCATCAG 3660  
Dd 3601 TGGAAACAAACCTTAGGAGATGGGCAGATGAGAGGGGACCATGATTTGAACCTGCATCAG 3660  
Oy 3661 ATAGCTTGGGGTCTTATGACCAATGTCGAGAGGGCCGAGGCTTATATGCTACAGCTACAG 3720  
Dd 3661 ATAGCTTGGGGTCTTATGACCAATGTCGAGAGGGCCGAGGCTTATATGCTACAGCTACAG 3720  
Oy 3721 TAATTTCTTCTCCAAACAGAGAACTTTCCAGAGATCAGGGGGATTCGGGCTCACCTTG 3780  
Dd 3721 TAATTTCTTCTCCAAACAGAGAACTTTCCAGAGATCAGGGGGATTCGGGCTCACCTTG 3780  
Oy 3781 ATGCTGTACAGATGGCCGTGGAGCTGAGAGTCAATGCTCAAGTGGCTCCATGATTAATA 3840  
Dd 3781 ATGCTGTACAGATGGCCGTGGAGCTGAGAGTCAATGCTCAAGTGGCTCCATGATTAATA 3840

QY	3841	TACAGACGATCCAGACCCAGAGAAAGCTGGGAGACTCTTCCATTGTGGGCACTATCTACCTTG	3900
Db	3841	TACAGACGATCCAGACCCAGAGAAAGCTGGGAGACTCTTCCATTGTGGGCACTATCTACCTTG	3900
QY	3901	ATTATTCAAGGGGATCCCTCAGATTATGCGCATCAAGCAGCATATGGAACCAAAATTATGT	3960
Db	3901	ATTATTCAAGGGGATCCCTCAGATTATGCGCATCAAGCAGCATATGGAACCAAAATTATGT	3960
QY	3961	TTTTCTGATCATAGCACAAGTATTAACAGGCAAAATCAAAGTATGAGAGAGCCTTGAAACAG	4020
Db	3961	TTTTCTGATCATAGCACAAGTATTAACAGGCAAAATCAAAGTATGAGAGAGCCTTGAAACAG	4020
QY	4021	CCCACTCCCGACCAAGCTGGGCGCTCTTCACAGCTGTACTGGGAGAAACTCAGAAAGTGG	4080
Db	4021	CCCACTCCCGACCAAGCTGGGCGCTCTTCACAGCTGTACTGGGAGAAACTCAGAAAGTGG	4080
QY	4081	ACACAGGCACAAATAAACGGAGGGGTGAAAGAGATGTTTCATTGAAGCCGAACCACTA	4140
Db	4081	ACACAGGCACAAATAAACGGAGGGGTGAAAGAGATGTTTCATTGAAGCCGAACCACTA	4140
QY	4141	GCCTAAGCTGTGTGACTACGGAAAGAAACCAACCGTCCCATGCTGCCATAGCTG	4200
Db	4141	GCCTAAGCTGTGTGACTACGGAAAGAAACCAACCGTCCCATGCTGCCATAGCTG	4200
QY	4201	TGGCATCAAGTACTACAAAGGGGCTTATTGCACGAAAGAGGGCAGTATGAGAGCCCC	4260
Db	4201	TGGCATCAAGTACTACAAAGGGGCTTATTGCACGAAAGAGGGCAGTATGAGAGCCCC	4260
QY	4261	CGCCACACCCTCCCGGCTACATTGGAATTCCTATTACTACTTCCAGAAAGGGCACTCCC	4320
Db	4261	CGCCACACCCTCCCGGCTACATTGGAATTCCTATTACTACTTCCAGAAAGGGCACTCCC	4320
QY	4321	ATCCAGCAGAGAAACGGCGGACTACAAACGTGAGCCCTTCAGAGANTCGCGATGTGTGCAC	4380
Db	4321	ATCCAGCAGAGAAACGGCGGACTACAAACGTGAGCCCTTCAGAGANTCGCGATGTGTGCAC	4380
QY	4381	GATCCTTCGACACAGCTGGGCTTTCATCCGTAAAGCAGCCAACTGGGGATCCCAACAGCA	4440
Db	4381	GATCCTTCGACACAGCTGGGCTTTCATCCGTAAAGCAGCCAACTGGGGATCCCAACAGCA	4440
QY	4441	GCAGGCTTGAAACAAACCTCAGTGGCATTAACCGAAGCATGTGCACCCGCGCTCGCCC	4500
Db	4441	GCAGGCTTGAAACAAACCTCAGTGGCATTAACCGAAGCATGTGCACCCGCGCTCGCCC	4500
QY	4501	CTTATCAGTCCCAAGGGTTCCTCCACCGAGAGGTGAAGATGAACAAGTTCTGTGCTGTTT	4560
Db	4501	CTTATCAGTCCCAAGGGTTCCTCCACCGAGAGGTGAAGATGAACAAGTTCTGTGCTGTTT	4560
QY	4561	GAGGCACAGACTTTTCTGGAAGCGAGCGAGGCCACTGAAAGGAGAGCACAGAAGACGT	4620
Db	4561	GAGGCACAGACTTTTCTGGAAGCGAGCGAGGCCACTGAAAGGAGAGCACAGAAGACGT	4620
QY	4621	CCTGAGCATTTGGAGCCTTGGAACTCACATTCGAGACGGGTGACCAAGTTGCTCCTTC	4680
Db	4621	CCTGAGCATTTGGAGCCTTGGAACTCACATTCGAGACGGGTGACCAAGTTGCTCCTTC	4680
QY	4681	CCTGCTTTAAAGCAGCATGGGGCTTCTCCGCCCTTCTCCCTTCCCTTTGCATGCGA	4740
Db	4681	CCTGCTTTAAAGCAGCATGGGGCTTCTCCGCCCTTCTCCCTTCTCCCTTTGCATGCGA	4740
QY	4741	AATACTGTGAAGAAATTCCTCGGCACATTTCAGACTTGTGTGCTGAAATGACAGTGC	4800
Db	4741	AATACTGTGAAGAAATTCCTCGGCACATTTCAGACTTGTGTGCTGAAATGACAGTGC	4800
QY	4801	AGCAATCTTCGAGCTCCCATCTGTGCTGCTGCCATACACAGTATCATTTCCAAATTC	4860
Db	4801	AGCAATCTTCGAGCTCCCATCTGTGCTGCTGCCATACACAGTATCATTTCCAAATTC	4860
QY	4861	CAGAGTCATCAACAACAAGTGAATTAAGTCTGGGCTCAGTCTTCATAGCTGCGTGAAGATTT	4920
Db	4861	CAGAGTCATCAACAACAAGTGAATTAAGTCTGGGCTCAGTCTTCATAGCTGCGTGAAGATTT	4920
QY	4921	TTTTTAATCTTCCTTTTGAATTTCAATCCAGTCTTAGACACTTGATCTATTGGGATATG	4980

Db	49 21	TTTTAAATCTCTCTTTTGGATTTCAATCAGCTTACGACTTGTATCTCATTTGGATATATG	4980
Qy	4981	ACAAAAGCTAGCATTGGAACACTGTTGGGGCCCTTAAACCACCAAGAGAACACAAAAGAAAA	5040
Db	4981	ACAAAAGCTAGCATTGGAACACTGTTGGGGCCCTTAAACCACCAAGAGAACACAAAAGAAAA	5040
Qy	5041	CAATGAATCTTTGAGTACAGTGGTGTCCACTGGTTTACAAATGCTCTCTTTAAAAA	5100
Db	5041	CAATGAATCTTTGAGTACAGTGGTGTCCACTGGTTTACAAATGCTCTCTTTAAAAA	5100
Qy	5101	AAAAATGAGTTTAAAGATTGTTGTCAGAGATAATATATCATTCATTAAATGATTACAG	5160
Db	5101	AAAAATGAGTTTAAAGATTGTTGTCAGAGATAATATATCATTCATTAAATGATTACAG	5160
Qy	5161	TATTAATTTTAAACCTTAAGTACGTTGCCAGCTTGGTTTCTGAAAACCAATATGCCGG	5220
Db	5161	TATTAATTTTAAACCTTAAGTACGTTGCCAGCTTGGTTTCTGAAAACCAATATGCCGG	5220
Qy	5221	ACAGGGTGGGCCACACCAAGAGAGGGAAGACCCTGTTGACCCCTGGCTTCCCATG	5280
Db	5221	ACAGGGTGGGCCACACCAAGAGAGGGAAGACCCTGTTGACCCCTGGCTTCCCATG	5280
Qy	5281	TCCCTGTGCTCTACCCCGGAACTGCCCCTATCCTGGAATATGAATGTTACCAATTAA	5340
Db	5281	TCCCTGTGCTCTACCCCGGAACTGCCCCTATCCTGGAATATGAATGTTACCAATTAA	5340
Qy	5341	TACCAAGACACCTCATCTGCTCCTTCCCAAGTACGATGGGGTCTTGTAAAACTGTTTG	5400
Db	5341	TACCAAGACACCTCATCTGCTCCTTCCCAAGTACGATGGGGTCTTGTAAAACTGTTTG	5400
Qy	5401	CACATGGCCAGGGAGAGGAACATAGAACCCCTGTGTCTGTCTGAGCCTTATGGAGCAGG	5460
Db	5401	CACATGGCCAGGGAGAGGAACATAGAACCCCTGTGTCTGTCTGAGCCTTATGGAGCAGG	5460
Qy	5461	ACGCTGTCATTTGGCGGATGTGTCTCTGCTCATGAGATGAGATGGCAACCCCATTTTAA	5520
Db	5461	ACGCTGTCATTTGGCGGATGTGTCTCTGCTCATGAGATGAGATGGCAACCCCATTTTAA	5520
Qy	5521	GTTAATATTCCTTGATTTTGTTAATTTAGAGGTAGGTTTGTGTTTTTGT	5580
Db	5521	GTTAATATTCCTTGATTTTGTTAATTTAGAGGTAGGTTTGTGTTTTTGT	5580
Qy	5581	TTTTTTTAAAGAAACATTTATACCTGGAATACCTGCAATGGAAGACAGCTTGGATGTT	5640
Db	5581	TTTTTTTAAAGAAACATTTATACCTGGAATACCTGCAATGGAAGACAGCTTGGATGTT	5640
Qy	5641	GGAGCTATGCGACCTGTTTATCTGCTCTTTCAGACACCTCCCTTTATGAATGGC	5700
Db	5641	GGAGCTATGCGACCTGTTTATCTGCTCTTTCAGACACAGCTCCCTTTATGAATGGC	5700
Qy	5701	ATTAGGAATTAACAGCCTTTTAAACGATGATAAAAATCAAAAACCTGTTTACACATGCC	5760
Db	5701	ATTAGGAATTAACAGCCTTTTAAACGATGATAAAAATCAAAAACCTGTTTACACATGCC	5760
Qy	5761	AGCCTTTCAGAGCAGGTATGTCACCAAGACTAACCACCAAGGCGCTTATATGAGCCTG	5820
Db	5761	AGCCTTTCAGAGCAGGTATGTCACCAAGACTAACCACCAAGGCGCTTATATGAGCCTG	5820
Qy	5821	CATATAGAGAAAGCCTTAGTGTAGCAACCATCTGCTACAGCTCTATTAACCTATAT	5880
Db	5821	CATATAGAGAAAGCCTTAGTGTAGCAACCATCTGCTACAGCTCTATTAACCTATAT	5880
Qy	5881	GACTGAATGACCCCTCCACTCTATTTTGTGTGTTTGTGACAGACTCCGGAAGAGGA	5940
Db	5881	GACTGAATGACCCCTCCACTCTATTTTGTGTGTTTGTGACAGACTCCGGAAGAGGA	5940
Qy	5941	AGCTGCCAATCTGATAGTACTCAAAATGTGAGAACTGCTGCTTGGATTTTTTTTCC	6000
Db	5941	AGCTGCCAATCTGATAGTACTCAAAATGTGAGAACTGCTGCTTGGATTTTTTTTCC	6000
Qy	6001	ATTAAATCAGCTGATCATATTGATCAGTAGATAAAGTAATATGCTTCAAAATTTTAAA	6066



Db 6001 ATTAAATTCAGCTGATCATATTGATCAGTAGATGAATTAACGTAAGTCAATTTTAA 6060  
Qy 6061 GTCGATTCAGCTGTTTTCACCTGATCAACAAATGATGCTTATTATTATTTCT 6120  
Db 6061 GTCGATTCAGCTGTTTTCACCTGATCAACAAATGATGCTTATTATTATTTCT 6120  
Qy 6121 CTCTGTCATCAGCAGCTTGTCTACTGCTTATTTACATTTGCAATTTGTAAT 6180  
Db 6121 CTCTGTCATCAGCAGCTTGTCTACTGCTTATTTACATTTGCAATTTGTAAT 6180  
Qy 6181 TTACATGTAATATGCAATTTTTCAGCTTTTATATATAGGCTAGGACCTCATGTGAT 6240  
Db 6181 TTACATGTAATATGCAATTTTTCAGCTTTTATATATAGGCTAGGACCTCATGTGAT 6240  
Qy 6241 ATAGAAAGACAGAAATCTGCTCTACCAAGTGCACAAATTTTCTAAGCATTAAGT 6300  
Db 6241 ATAGAAAGACAGAAATCTGCTCTACCAAGTGCACAAATTTTCTAAGCATTAAGT 6300  
Qy 6301 AATTGTAGAACATAGACATGCTAATCTCAGTTCGCTCTGTGATGTCAGTGCAGAAATGTA 6360  
Db 6301 AATTGTAGAACATAGACATGCTAATCTCAGTTCGCTCTGTGATGTCAGTGCAGAAATGTA 6360  
Qy 6361 CAATTACAGTGTGATTTCTCATCTTGTGATTTACTTGTACTGTATGCTTTTGA 6420  
Db 6361 CAATTACAGTGTGATTTCTCATCTTGTGATTTACTTGTACTGTATGCTTTTGA 6420  
Qy 6421 AGACATTTGGTGGAGTCTGATCCCTTTGATTTTAAATACAAATTTGATGATTTGGT 6480  
Db 6421 AGACATTTGGTGGAGTCTGATCCCTTTGATTTTAAATACAAATTTGATGATTTGGT 6480  
Qy 6481 TATATTTTGTGGAAGATGATGAAATGTAATGTAATGTTATGCTTTCATCCAGTTTGA 6540  
Db 6481 TATATTTTGTGGAAGATGATGAAATGTAATGTAATGTTATGCTTTCATCCAGTTTGA 6540  
Qy 6541 CAAGCTGGAATAATTAATTAATTAACAT 6568  
Db 6541 CAAGCTGGAATAATTAATTAATTAACAT 6568

RESULT 2  
ABN95037  
ID ABN95037 standard; DNA; 6568 BP.  
AC ABN95037;  
DT 13-AUG-2002 (first entry)  
DE Gene #1535 used to diagnose liver cancer.  
XX Gene: liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
KM metastatic liver tumour; cytostatic; expression profile; disease state;  
KM disease progression; drug toxicity; drug efficacy; drug metabolism.  
OS Homo sapiens.  
XX WO200229103-A2.  
XX 11-APR-2002.  
XX 02-OCT-2001; 2001WO-US30589.  
XX 02-OCT-2001; 2000US-237054P.  
XX (GENE-) GENE LOGIC INC.  
XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;  
PI MPI; 2002-426119/45.  
XX Diagnosing and detecting the progression of liver cancer,  
PT hepatocellular carcinoma or metastatic liver tumor in a patient,  
PT involves detecting the level of expression of two or more genes in a  
PT liver tissue sample

XX  
PS Claim 1: SEQ ID NO 1535; 298bp; English.  
CC The invention relates to a novel method for diagnosing and detecting the  
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
CC tumor in a patient, and differentiating metastatic liver cancer from  
CC hepatocellular carcinoma in a patient, involving detecting the level of  
CC expression of two or more genes represented in ABN95037-ABN97455 in a  
CC tissue sample. The method of the invention has hepatotropic, and  
CC cytosolic activity. The method is useful for diagnosing and detecting  
CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
CC liver carcinoma in a patient. The method is useful for identifying  
CC expression profiles which serve as useful diagnostic markers as well as  
CC markers that can be used to monitor disease states, disease progression,  
CC drug toxicity, drug efficacy and drug metabolism.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pat\_sequences.  
XX  
SQ Sequence 6568 BP; 1974 A; 1400 C; 1463 G; 1731 T; 0 other:

Query Match 100.0%; Score 6568; DB 24; Length 6568;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGGCCATCGTGAGAGATTGCTACATGATGTAATTCAGTTCAGCATATGTTCTTCA 60  
Db 1 CTGGCCATCGTGAGAGATTGCTACATGATGTAATTCAGTTCAGCATATGTTCTTCA 60  
Qy 61 TTATGAACCACTAGCAATCCAGCTAACCATGGAGTTATGGGCCACAGAGAAACACT 120  
Db 61 TTATGAACCACTAGCAATCCAGCTAACCATGGAGTTATGGGCCACAGAGAAACACT 120  
Qy 121 CACTTCCTGCAGATTTTCACAAACCTGCATTTACTGACAGTCTCCACACAGGTGACC 180  
Db 121 CACTTCCTGCAGATTTTCACAAACCTGCATTTACTGACAGTCTCCACACAGGTGACC 180  
Qy 181 AGCTTCTCTTACCCATTCGAGATGTACACTGATGATGTTCTGGAGAGACAGCTTT 240  
Db 181 AGCTTCTCTTACCCATTCGAGATGTACACTGATGATGTTCTGGAGAGACAGCTTT 240  
Qy 241 CTGATATCTACACAGCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
Db 241 CTGATATCTACACAGCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
Qy 301 AAACAGCAGTGTGATTTCCGAAGACGAGCATGAGAGAGAGAGAGAGAGAGAGATC 360  
Db 301 AAACAGCAGTGTGATTTCCGAAGACGAGCATGAGAGAGAGAGAGAGAGAGAGATC 360  
Qy 361 CTCTGATGACGAGGAGCATTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
Db 361 CTCTGATGACGAGGAGCATTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
Qy 421 ATGATGACATTTGAACAACCTGTAATTTATGACACAGTTCCTGCTTTGCCAATATGA 480  
Db 421 ATGATGACATTTGAACAACCTGTAATTTATGACACAGTTCCTGCTTTGCCAATATGA 480  
Qy 481 CAATGTCAGTGAAGCGAGAACTCTGTGCTGTGATGCTTTCGAGAGTGTGAGAGAGAG 540  
Db 481 CAATGTCAGTGAAGCGAGAACTCTGTGCTGTGATGCTTTCGAGAGTGTGAGAGAGAG 540  
Qy 541 GGAACATAGTGTAAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
Db 541 GGAACATAGTGTAAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
Qy 601 CTGTGAGAGTGAATTTCCAGATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
Db 601 CTGTGAGAGTGAATTTCCAGATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
Qy 661 GTGTCTCTCTACCATGAG 720  
Db 661 GTGTCTCTCTACCATGAG 720

Qy	721	ACTGCCACTTTTGTCGATATAGCCAGCAAGATTA	CTGCGCGTATTCTCATATCAAGTACAAA	780
Db	721	ACTGCCAGTTTGTCGATATAGCCAGCAAGATTA	CTGCGCGTATTCTCATATCAAGTACAAA	780
Qy	781	AGACATGCAAAAAGTTGAAGAGGAAGATTTGTTAT	TGGTGAAGAACACCGAGAAC	840
Db	781	AGAACATGCAAAAAGTTGAAGAGGAAGATTTGTTAT	TGGTGAAGAACACCGAGAAC	840
Qy	841	TTGATCGAACTGGAACAAGAAAGGACACATTTGTCAT	CAAGGGTACCTCAGAAAGGTTAA	900
Db	841	TTGATCGAACTGGAACAAGAAAGGACACATTTGTCAT	CAAGGGTACCTCAGAAAGGTTAA	900
Qy	901	CAATGATATTTGGTGGAGAGACATTTCAGTATGATAT	CAACAAATTCATATAAGACTTTCTGT	960
Db	901	CAATGATATTTGGTGGAGAGACATTTCAGTATGATAT	CAACAAATTCATATAAGACTTTCTGT	960
Qy	961	TGACCTATATGAACTTTTCTTTCTTACCCCAATGGA	AGTGGGCAAAAAAGTTATTTGAGTGGT	1020
Db	961	TGACCTATATGAACTTTTCTTTCTTACCCCAATGGA	AGTGGGCAAAAAAGTTATTTGAGTGGT	1020
Qy	1021	TTATATGACCCGAGCCTCAGGGATTAAGGTTACAC	CGGGTACTATTATTGGGGTAAATATAC	1080
Db	1021	TTATATGACCCGAGCCTCAGGGATTAAGGTTACAC	CGGGTACTATTATTGGGGTAAATATAC	1080
Qy	1081	ACCTTCATATGACTTTTGAAGAGATCTCGCAATGAC	TATCTGATTTTATAGAAATTTGAAAAACA	1140
Db	1081	ACCTTCATATGACTTTTGAAGAGATCTCGCAATGAC	TATCTGATTTTATAGAAATTTGAAAAACA	1140
Qy	1141	ATCTGAAAGAGAGAAATATGGGTGACACCTTAAG	CGCTGTTGAATATCGCGTGTCTGCTA	1200
Db	1141	ATCTGAAAGAGAGAAATATGGGTGACACCTTAAG	CGCTGTTGAATATCGCGTGTCTGCTA	1200
Qy	1201	AACCAAAAAGAAATATGATGACGTTTAACAAAAC	CAATCCCGAGAGCTCCTTTGGCCTTTTA	1260
Db	1201	AACCAAAAAGAAATATGATGATGACGTTTAACAAAAC	CAATCCCGAGAGCTCCTTTGGCCTTTTA	1260
Qy	1261	TCTTACTTTGGAGGCTCTGAGAAAGGATTTGGAAT	CTTTGTTGACAGTGTAGATTCAAGGTA	1320
Db	1261	TCTTACTTTGGAGGCTCTGAGAAAGGATTTGGAAT	CTTTGTTGACAGTGTAGATTCAAGGTA	1320
Qy	1321	GCAAAAGCACTGAAAGCAAGGCTTAAACGGGGGAT	CATGATATTAGAAAGTAAATGGCCAAA	1380
Db	1321	GCAAAAGCACTGAAAGCAAGGCTTAAACGGGGGAT	CATGATATTAGAAAGTAAATGGCCAAA	1380
Qy	1381	ACTTTGAAAACATTTGACCTGTCAAAAAGCTATG	GAATTTCTTAGAAATTAACACACTTTTAT	1440
Db	1381	ACTTTGAAAACATTTGACCTGTCAAAAAGCTATG	GAATTTCTTAGAAATTAACACACTTTTAT	1440
Qy	1441	CTATCACTGTGAAACCAATTTATTTGTATTTAAGA	CACTTCTAACAAAGTTGTTCAGAAAG	1500
Db	1441	CTATCACTGTGAAACCAATTTATTTGTATTTAAGA	CACTTCTAACAAAGTTGTTCAGAAAG	1500
Qy	1501	AGAAAGAAATGTGTGCCCCCACCCTTCTTAAAT	TGGTGACATTTAAAGGCCAATGTCCT	1560
Db	1501	AGAAAGAAATGTGTGCCCCCACCCTTCTTAAAT	TGGTGACATTTAAAGGCCAATGTCCT	1560
Qy	1561	ACTCGATTTCCAGATCTTGTCTGTATATGATAGAC	AGGTGATATGACACTTGAAAAAGTGTACA	1620
Db	1561	ACTCGATTTCCAGATCTTGTCTGTATATGATAGAC	AGGTGATATGACACTTGAAAAAGTGTACA	1620
Qy	1621	AAAAAAGTAAGCCACAACACTGTGGAGAGAAAGAA	CAAGCTGAAAAAGATATCTGCACAGA	1680
Db	1621	AAAAAAGTAAGCCACAACACTGTGGAGAGAAAGAA	CAAGCTGAAAAAGATATCTGCACAGA	1680
Qy	1681	CTCGGATCAGTATCTTGGCACAGAAACATATGATAT	TGGGATGGTGTGACGTCTGCTAAG	1740
Db	1681	CTCGGATCAGTATCTTGGCACAGAAACATATGATAT	TGGGATGGTGTGACGTCTGCTAAG	1740
Qy	1741	ATGACAGATATGATAGATTAAAGGACAGACAAGACA	TATCCCAACTGCAATTTGGCTGTCAAGT	1800
Db	1741	ATGACAGATATGATAGATTAAAGGACAGACAAGACA	TATCCCAACTGCAATTTGGCTGTCAAGT	1800
Qy	1801	GAACTTTTCATCCAGTAATCCTGATTTATTGCA	GTCAATCATCATTCGACTTTTACACTTCA	1860

Db	1801	GAACCTTTACATCCCAAGTAACTCTGATTTATTCGATCCATCATTCGATTTTAACTTCA	1860
Qy	1861	GGCTCACTCCCTACCTTGGCCAGATCAAGTCGAAGGGTTTTAAAGCTGATCAGCAAAACC	1920
Db	1861	GTGCTAATCTGACTTGGCCAGATCAAAAGTGCTAAAGGGTTTTAAAGCTGATCAGCAAAACC	1920
Qy	1921	GCTACATCATGATCATAGTAAGCACTACACGCAAAAGAAAGTGGTCAATTGAGCTATCAGG	1980
Db	1921	GCTACATCATGATCATAGTAAGCACTACACGCAAAAGAAAGTGGTCAATTGAGCTATCAGG	1980
Qy	1981	AGTTGGCTTATCTGCGACCCCGAGTCAATATTCACATATGTAGAGTCTCTGTCCACCCTG	2040
Db	1981	AGTTTCTCTTACTGCGACCCCGAGTCAATATTCACATATGTAGAGTCTCTGTCCACCCTG	2040
Qy	2041	AGGAGCTAAATCAAAACAAGAAGCTTCCAGATCAGCTTTCCAACTTGGCAGCAACAATAC	2100
Db	2041	AGGAGCTAAATCAAAACAAGAAGCTTCCAGATCAGCTTTCCAACTTGGCAGCAACAATAC	2100
Qy	2101	AACGTGCTGAAGGTTATTTCTGAAAAAACAACATGGAAACAACAACCTCTTTTTCAGATG	2160
Db	2101	AACGTGCTGAAGGTTATTTCTGAAAAAACAACATGGAAACAACAACCTCTTTTTCAGATG	2160
Qy	2161	AAGATCCTCAGSAGTTGTTGAGAGAGTCAAAATTTCCCTCCTCAGCTCAGCAGCTGTGG	2220
Db	2161	AAGATGCTCAGSAGTTGTTGAGAGAGTCAAAATTTCCCTCCTCAGCTCAGCAGCTGTGG	2220
Qy	2221	AACTTGCACACAGCTCTCATGCGGAAATTTGAACCTCTTTGGCAACATTTGAACCTACTG	2280
Db	2221	AACTTGCACACAGCTCTCATGCGGAAATTTGAACCTCTTTGGCAACATTTGAACCTACTG	2280
Qy	2281	AATATATAGATGATTTATTTAACTCAGATCAAAAACACAGCTGTGGCCAACTGGAAGAT	2340
Db	2281	AATATATAGATGATTTATTTAACTCAGATCAAAAACACAGCTGTGGCCAACTGGAAGAT	2340
Qy	2341	TTGAGAAGATCATTTAACGAGAAACATTTTGGGTAGACATCTGAAATTTCCAGAGAAACA	2400
Db	2341	TTGAGAAGATCATTTAACGAGAAACATTTTGGGTAGACATCTGAAATTTCCAGAGAAACA	2400
Qy	2401	ACCGCTGAGAGAGATGAATCATTTAAGCATTTTCATTCATCAAGATAGCAGTCACTGTAGGG	2460
Db	2401	ACCGCTGAGAGAGATGAATCATTTAAGCATTTTCATTCATCAAGATAGCAGTCACTGTAGGG	2460
Qy	2461	AATGCAAGATTTTAACTCAATGTTTGCATCATCACTGAGGCTTAAACCTGGCACAGTGG	2520
Db	2461	AATGCAAGATTTTAACTCAATGTTTGCATCATCACTGAGGCTTAAACCTGGCACAGTGG	2520
Qy	2521	CAAGACGCGCAACACCTGGGAGAACTTCCATATAAATACGAAAAACTATTTTCAGATC	2580
Db	2521	CAAGACGCGCAACACCTGGGAGAACTTCCATATAAATACGAAAAACTATTTTCAGATC	2580
Qy	2581	TCGAAGACCTGTTGATCTTCCAGAAACATGCGAAATATCTGAATGGTCTCAATATGTC	2640
Db	2581	TCGAAGACCTGTTGATCTTCCAGAAACATGCGAAATATCTGAATGGTCTCAATATGTC	2640
Qy	2641	AAAATCTCAACCTGCCATATCCCTCTATTTCCAGTTATCAAAAAGATCTCACCTTCC	2700
Db	2641	AAAATCTCAACCTGCCATATCCCTCTATTTCCAGTTATCAAAAAGATCTCACCTTCC	2700
Qy	2701	TTTCAGAGGAATGACTCAAAAGTAGACGGGCTGATCAATTTTGAAGAGTAAAGATGA	2760
Db	2701	TTTCAGAGGAATGACTCAAAAGTAGACGGGCTGATCAATTTTGAAGAGTAAAGATGA	2760
Qy	2761	TTTGCAAAAGAATTCGTCACGTTGGCCGAATGGCTCGTAACATGAGACCTGCCCCCTCA	2820
Db	2761	TTTGCAAAAGAATTCGTCACGTTGGCCGAATGGCTCGTAACATGAGACCTGCCCCCTCA	2820
Qy	2821	TGTTCAAGACTCGAAGAAAGAAATGGCGAGATTTGGGGTCTCTCAGCCAGGGTAGTACAA	2880
Db	2821	TGTTCAAGACTCGAAGAAAGAAATGGCGAGATTTGGGGTCTCTCAGCCAGGGTAGTACAA	2880
Qy	2881	ATGCAACAGTGTAGATGTGCTCAGACAGGTGCTATAAAAACGGGGTACGTGTGTT	2940

Dh 2881 ATGCAAGTGTCTAGATGTTGCTCAGACAGGTGTCATAAAAAGGGGTACGTCTAGTT 2940  
Qy 2941 CCTTTCATATGCCAAAAGCTTTATGAAGATGCCCAAATGGCTCGAAAAGTGAAGAGT 3000  
Dh 2941 CCTTTCATATGCCAAAAGCTTTATGAAGATGCCCAAATGGCTCGAAAAGTGAAGAGT 3000  
Qy 3001 ACCTTTCAATTTGGAGCTGAATGAGCAGAGAGTCTTTCAGACATTTATCTCTGCAGT 3060  
Dh 3001 ACCTTTCAATTTGGAGCTGAATGAGCAGAGAGTCTTTCAGACATTTATCTCTGCAGT 3060  
Qy 3061 GTGAGCCAGCAACACACATGTGCTTAAGAACTCTGTGACAAAAGCTGTCAATTCG 3120  
Dh 3061 GTGAGCCAGCAACACACATGTGCTTAAGAACTCTGTGACAAAAGCTGTCAATTCG 3120  
Qy 3121 AGACCTTCAGTGTGCTCAAGGGGCTACACAGAAAAGCTCACTCCCTGCGACAGC 3180  
Dh 3121 AGACCTTCAGTGTGCTCAAGGGGCTACACAGAAAAGCTCACTCCCTGCGACAGC 3180  
Qy 3181 CCCAGCAGAGCCACACACAGACATATAATCAACAGGGAGTACAGAGTTCGCCGTGT 3240  
Dh 3181 CCCAGCAGAGCCACACACAGACATATAATCAACAGGGAGTACAGAGTTCGCCGTGT 3240  
Qy 3241 CCTTTATCTTCAAGGAGAAAGTCCCGTAAGAGTCTCCACCTTTTGGCATAACT 3300  
Dh 3241 CCTTTATCTTCAAGGAGAAAGTCCCGTAAGAGTCTCCACCTTTTGGCATAACT 3300  
Qy 3301 CTCACAGCTTTAAAAAATTTCTTTCTGTGTGAAGAAAGATTGGAAGTCA 3360  
Dh 3301 CTCACAGCTTTAAAAAATTTCTTTCTGTGTGAAGAAAGATTGGAAGTCA 3360  
Qy 3361 AGAAGAGGCTGAAGATACATATCAATATGATCTGCGACCTTTCTCTCTCTACTT 3420  
Dh 3361 AGAAGAGGCTGAAGATACATATCAATATGATCTGCGACCTTTCTCTCTCTACTT 3420  
Qy 3421 CTCACAGATTTCTCCAGAGAAAGGCTATACTTGGCTCCAGTGTGATGTAATTT 3480  
Dh 3421 CTCACAGATTTCTCCAGAGAAAGGCTATACTTGGCTCCAGTGTGATGTAATTT 3480  
Qy 3481 TTTAGATTTGGTTCAGAGTAATAATTTCTTACAGTTCAGATTTAGCAATTCGCTT 3540  
Dh 3481 TTTAGATTTGGTTCAGAGTAATAATTTCTTACAGTTCAGATTTAGCAATTCGCTT 3540  
Qy 3541 TTTAGTCTAGTCCAGTCTCACTGCGACGATGAGAGCGCCAGAGGCAATTTCTGACATCG 3600  
Dh 3541 TTTAGTCTAGTCCAGTCTCACTGCGACGATGAGAGCGCCAGAGGCAATTTCTGACATCG 3600  
Qy 3601 TGGAAACAAACCTAGGGATGGGCGAGATGGAGCGGACCATGATGAACTGATCAGT 3660  
Dh 3601 TGGAAACAAACCTAGGGATGGGCGAGATGGAGCGGACCATGATGAACTGATCAGT 3660  
Qy 3661 ATACTTGGGGTCCATGACATGATCCGAGGGCCGAGGCTTATATGCTACACTACAG 3720  
Dh 3661 ATACTTGGGGTCCATGACATGATCCGAGGGCCGAGGCTTATATGCTACACTACAG 3720  
Qy 3721 TAAATTTCTTCCAGAGCAGAGAACTTTCCAGAGTCAAGGGAGTCCGCTCACTTG 3780  
Dh 3721 TAAATTTCTTCCAGAGCAGAGAACTTTCCAGAGTCAAGGGAGTCCGCTCACTTG 3780  
Qy 3781 ATGCTGCTGACATGGCCGTGGAGCTGAGCTATGCTCAAGTGGCTCCCATGATAATA 3840  
Dh 3781 ATGCTGCTGACATGGCCGTGGAGCTGAGCTATGCTCAAGTGGCTCCCATGATAATA 3840  
Qy 3841 TACAGAGATTCACACACAGAGAAAGTGGGAGCTTTCCATTTGGGGATACTCACTTG 3900  
Dh 3841 TACAGAGATTCACACACAGAGAAAGTGGGAGCTTTCCATTTGGGGATACTCACTTG 3900  
Qy 3901 ATTATTCAGGGGATCTGACAGTTATGGGATCAAGCAGCATATGACCAAAATATGT 3960  
Dh 3901 ATTATTCAGGGGATCTGACAGTTATGGGATCAAGCAGCATATGACCAAAATATGT 3960  
Qy 3961 TTTCTGATATGACAAAGATATTAACAGGCAAAATCAAAAGTAGAGAGCTTGAACAG 4020  
Dh 3961 TTTCTGATATGACAAAGATATTAACAGGCAAAATCAAAAGTAGAGAGCTTGAACAG 4020  
Qy 4021 CCCAGTCCCGAGCAAGCTGGGCGTCTTCCACAGGTTACTGGGGAGAAAGACTCAGAGGTG 4080  
Dh 4021 CCCAGTCCCGAGCAAGCTGGGCGTCTTCCACAGGTTACTGGGGAGAAAGACTCAGAGGTG 4080  
Qy 4081 ACACAGCACAATTAAGCGAGGGGTGAAGAGATGTTTCCATTGAAGCCGAAGACGTA 4140  
Dh 4081 ACACAGCACAATTAAGCGAGGGGTGAAGAGATGTTTCCATTGAAGCCGAAGACGTA 4140  
Qy 4141 GCTTAAGTCTGTACTATGCAAGAAACCAAGCTGTCCCATGTCCCATATAGCTG 4200  
Dh 4141 GCTTAAGTCTGTACTATGCAAGAAACCAAGCTGTCCCATGTCCCATATAGCTG 4200  
Qy 4201 TGGCATCAAGTACTACAAAGGGGCTCATGTCAGAAAGAGGGAGGTATGAGAGCC 4260  
Dh 4201 TGGCATCAAGTACTACAAAGGGGCTCATGTCAGAAAGAGGGAGGTATGAGAGCC 4260  
Qy 4261 CGCCACCCCTCCCGGCTACATTTGGAATTTCCATTACTGACTTTCCAGAAAGGCACTCC 4320  
Dh 4261 CGCCACCCCTCCCGGCTACATTTGGAATTTCCATTACTGACTTTCCAGAAAGGCACTCC 4320  
Qy 4321 ATCCAGCCAGAAACCGCGGACTATCAAGTGGCCCTTCAAGATGCGGATGTCGAC 4380  
Dh 4321 ATCCAGCCAGAAACCGCGGACTATCAAGTGGCCCTTCAAGATGCGGATGTCGAC 4380  
Qy 4381 GATCCTCCGACACAGCTGGGCTTCATCCCTACAGACACCCACATGGGATCCACACGA 4440  
Dh 4381 GATCCTCCGACACAGCTGGGCTTCATCCCTACAGACACCCACATGGGATCCACACGA 4440  
Qy 4441 GCAGGCTGTGAACAAACCTAGTGGCATTAACAGAAAGATCGACCGCGCTCGCC 4500  
Dh 4441 GCAGGCTGTGAACAAACCTAGTGGCATTAACAGAAAGATCGACCGCGCTCGCC 4500  
Qy 4501 CTATACAGTCCCAAGGTTTTCACCGAGAGAGATGAAGATGAACAGTTCTGCTGTT 4560  
Dh 4501 CTATACAGTCCCAAGGTTTTCACCGAGAGAGATGAAGATGAACAGTTCTGCTGTT 4560  
Qy 4561 GAGGACAGACTTTTCTGGAAGCAGAGGAGGACCCCTGAAAGGAGGACAAAGAAAGAGT 4620  
Dh 4561 GAGGACAGACTTTTCTGGAAGCAGAGGAGGACCCCTGAAAGGAGGACAAAGAAAGAGT 4620  
Qy 4621 CCTGAGCATTTGAGACCTTGAACATCAATTTGAGAGAGGTGAGACATTTGCTCTTC 4680  
Dh 4621 CCTGAGCATTTGAGACCTTGAACATCAATTTGAGAGAGGTGAGACATTTGCTCTTC 4680  
Qy 4681 CCTGCTTTAAAGACAGATGGGCTTCTTCTCCCTTTCTCTTCCCTTTGCAATGGA 4740  
Dh 4681 CCTGCTTTAAAGACAGATGGGCTTCTTCTCCCTTTCTCTTCCCTTTGCAATGGA 4740  
Qy 4741 AATCTGTGAAGAAATTTCCCGGACATTTTCAGACTTTGCTTGAAGTGAACAGTGC 4800  
Dh 4741 AATCTGTGAAGAAATTTCCCGGACATTTTCAGACTTTGCTTGAAGTGAACAGTGC 4800  
Qy 4801 AGCATCTTTCAGGCTCCCACTGTTGCTGCTGCGCATACACAGATATCATTTCCAAATTC 4860  
Dh 4801 AGCATCTTTCAGGCTCCCACTGTTGCTGCTGCGCATACACAGATATCATTTCCAAATTC 4860  
Qy 4861 CAAGATCATCAACACAAATATTCACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4920  
Dh 4861 CAAGATCATCAACACAAATATTCACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4920  
Qy 4921 TTTTAAATCTTCTTTTGAATTTCAATCCAGTCTTCAAGATTTGATTTGATTTGATTTG 4980  
Dh 4921 TTTTAAATCTTCTTTTGAATTTCAATCCAGTCTTCAAGATTTGATTTGATTTGATTTG 4980  
Qy 4981 AGAAAGCTAGACATTTGAATTTGAGGCTTTTAACCCACCAAGAGAGACAAAGAAAA 5040  
Dh 4981 AGAAAGCTAGACATTTGAATTTGAGGCTTTTAACCCACCAAGAGAGACAAAGAAAA 5040  
Qy 5041 CAATGAATTCCTTTGATGATGATGCTTCCATCTTGTATTAAGTCTCTTTTAAAAA 5100  
Dh 5041 CAATGAATTCCTTTGATGATGATGCTTCCATCTTGTATTAAGTCTCTTTTAAAAA 5100

QY 5101 AAAAAATGATTTAAAGATTTTGTTCACAGAGTAATATATATCCATTATATGATACAG 5160  
|||||  
Db 5101 AAAAAATGATTTAAAGATTTTGTTCACAGAGTAATATATATCCATTATATGATACAG 5160  
QY 5161 TATTTATTTTAAACCTTAAGTGGGTTGCGACGCTGTTTCGAAAAACCAATATATGCCG 5220  
|||||  
Db 5161 TATTTATTTTAAACCTTAAGTGGGTTGCGACGCTGTTTCGAAAAACCAATATATGCCG 5220  
QY 5221 ACAGGGTGTGGCCACACCAAGAAAGACGGGAAGACCTGGCTTGACACCTGGCTCCATG 5280  
|||||  
Db 5221 ACAGGGTGTGGCCACACCAAGAAAGACGGGAAGACCTGGCTTGACACCTGGCTCCATG 5280  
QY 5281 TCCCTTCGTGCTCACCCCGGAGTCCCTATCTCTGGAAGTATGAATGTAGCCATTTAA 5340  
|||||  
Db 5281 TCCCTTCGTGCTCACCCCGGAGTCCCTATCTCTGGAAGTATGAATGTAGCCATTTAA 5340  
QY 5341 TACCAAGACACCTCATCTGCTCTCTCCAGTGCATGGGTTCTCTGTAACACTGTTTG 5400  
|||||  
Db 5341 TACCAAGACACCTCATCTGCTCTCTCCAGTGCATGGGTTCTCTGTAACACTGTTTG 5400  
QY 5401 CACATGGCCAGGGGAGGAGAACTAGAACCCCTTGCTGCTGCTGAGCCCTTATGAGCAGG 5460  
|||||  
Db 5401 CACATGGCCAGGGGAGGAGAACTAGAACCCCTTGCTGCTGCTGAGCCCTTATGAGCAGG 5460  
QY 5461 AGCGTGTCAATGGCGGATGTGCTGCTCATTTGAGATGATGGAAACCCCATTTTAA 5520  
|||||  
Db 5461 AGCGTGTCAATGGCGGATGTGCTGCTCATTTGAGATGATGGAAACCCCATTTTAA 5520  
QY 5521 GTTATATTTCTTGATTTTGTATTTAGAGGTAGGTTTGTGTTTTTGT 5580  
|||||  
Db 5521 GTTATATTTCTTGATTTTGTATTTAGAGGTAGGTTTGTGTTTTTGT 5580  
QY 5581 TTTTATTAAGAGAAACATTTATATAGTATGATGACATGCGTAAAGACACTGGGATGTT 5640  
|||||  
Db 5581 TTTTATTAAGAGAAACATTTATATAGTATGATGACATGCGTAAAGACACTGGGATGTT 5640  
QY 5641 GGAGCTAATGCGCAGCTGTTTATCTGCTTTCAAGACAGCCTCCCTTATTTGAATGGC 5700  
|||||  
Db 5641 GGAGCTAATGCGCAGCTGTTTATCTGCTTTCAAGACAGCCTCCCTTATTTGAATGGC 5700  
QY 5701 ATTAGGAATTAACAGCCCTTTAAGCTGATTAAGATCAAAAACCTGGTTAGACATGCG 5760  
|||||  
Db 5701 ATTAGGAATTAACAGCCCTTTAAGCTGATTAAGATCAAAAACCTGGTTAGACATGCG 5760  
QY 5761 AGGCTTGAAGGAGGAGTGTAGTACCAAGACTAACCTCAAGTGGCTTTATGAGACGTG 5820  
|||||  
Db 5761 AGGCTTGAAGGAGGAGTGTAGTACCAAGACTAACCTCAAGTGGCTTTATGAGACGTG 5820  
QY 5821 CATATAGAGAGGCTTAAGTGTAGCAACCATCTGTCACAGCTGATTTAACCTATAT 5880  
|||||  
Db 5821 CATATAGAGAGGCTTAAGTGTAGCAACCATCTGTCACAGCTGATTTAACCTATAT 5880  
QY 5881 GACTGAATATACCCCTTCACACTATTTTGTGTGTTTGGCAGACACTCGGAAAGTGA 5940  
|||||  
Db 5881 GACTGAATATACCCCTTCACACTATTTTGTGTGTTTGGCAGACACTCGGAAAGTGA 5940  
QY 5941 AGGCTGCAATGTAGTACTCAATGTAGAGAACTGCTGCTGTTGATTTTTTTTCC 6000  
|||||  
Db 5941 AGGCTGCAATGTAGTACTCAATGTAGAGAACTGCTGCTGTTGATTTTTTTTCC 6000  
QY 6001 ATTAATATTCAGCTGATCATATTCATCATAGATTAACGTAATATGCTTCAAAATTTTAA 6060  
|||||  
Db 6001 ATTAATATTCAGCTGATCATATTCATCATAGATTAACGTAATATGCTTCAAAATTTTAA 6060  
QY 6061 GTGGAATTTGAGTGTTTTTCACGTATCAAAACATGTGCTTATATTAATTTCC 6120  
|||||  
Db 6061 GTGGAATTTGAGTGTTTTTCACGTATCAAAACATGTGCTTATATTAATTTCC 6120  
QY 6121 CTTCGTATCATGATTTGCTACTGCTTATTTACATTTGCTCAATTTATGATTTGTAAT 6180  
|||||  
Db 6121 CTTCGTATCATGATTTGCTACTGCTTATTTACATTTGCTCAATTTATGATTTGTAAT 6180  
QY 6181 TTACATGTAAATATGATTAATTTTGGCAGTTTATTTATATAGGCTATGACCTCATGTGCAT 6240  
|||||

Db 6181 TTACATGTAAATATGATTAATTTTGGCAGTTTATTTATATAGGCTATGACCTCATGTGCAT 6240  
|||||  
QY 6241 ATGAGAAAGACAGAAATCTACCTTACCACAGTTGCACAATGTATTTATAGCATTTAACT 6300  
|||||  
Db 6241 ATGAGAAAGACAGAAATCTACCTTACCACAGTTGCACAATGTATTTATAGCATTTAACT 6300  
QY 6301 AATTGTGAACATATGAGACTGCTATCTCACTTGGCTCTGTGATGTCAAGTGCAGAAATGTA 6360  
|||||  
Db 6301 AATTGTGAACATATGAGACTGCTATCTCACTTGGCTCTGTGATGTCAAGTGCAGAAATGTA 6360  
QY 6361 CAATTAAGTGTGATTTCTCATATCTTTTGTATCTATCTTGTACCTGTATGCTCTTTTGA 6420  
|||||  
Db 6361 CAATTAAGTGTGATTTCTCATATCTTTTGTATCTATCTTGTACCTGTATGCTCTTTTGA 6420  
QY 6421 AGACATTTGTGAGAGTGTATCCCTTTGTATTTTATATCATTAATTTGATATTTGGT 6480  
|||||  
Db 6421 AGACATTTGTGAGAGTGTATCCCTTTGTATTTTATATCAATTAATTTGATATTTGGT 6480  
QY 6481 TATATTTTGTGTAAGATGGTGAATGTACTATGTTATGCTTCTATCATCCAGTTTGT 6540  
|||||  
Db 6481 TATATTTTGTGTAAGATGGTGAATGTACTATGTTATGCTTCTATCATCCAGTTTGT 6540  
QY 6541 CAAGCTGGAATAATTAATTAATTAACAT 6568  
|||||  
Db 6541 CAAGCTGGAATAATTAATTAATTAACAT 6568  
RESULT 3  
AAV84550  
ID AAV84550 standard; DNA; 2890 BP.  
XX  
AC AAV84550;  
XX  
DT 01-MAR-1999 (first entry)  
XX  
DE Human secreted protein gene 140 clone HE26720.  
XX  
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukemia;  
developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
conjunctive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
OS Homo sapiens.  
XX  
PN W09854963-A2.  
XX  
PD 10-DEC-1998.  
XX  
PE 04-JUN-1998; 98WO-US11422.  
XX  
PR 18-DEC-1997; 97US-0070923.  
PR 06-JUN-1997; 97US-0048877.  
PR 06-JUN-1997; 97US-0048881.  
PR 06-JUN-1997; 97US-0048884.  
PR 06-JUN-1997; 97US-0048893.  
PR 06-JUN-1997; 97US-0048896.  
PR 06-JUN-1997; 97US-0048899.  
PR 06-JUN-1997; 97US-0048915.  
PR 06-JUN-1997; 97US-0048949.  
PR 06-JUN-1997; 97US-0048964.  
PR 06-JUN-1997; 97US-0048972.  
PR 06-JUN-1997; 97US-0049020.  
PR 06-JUN-1997; 97US-0049375.  
PR 05-SEP-1997; 97US-0057628.  
PR 05-SEP-1997; 97US-0057635.  
PR 05-SEP-1997; 97US-0057644.  
PR 05-SEP-1997; 97US-0057647.  
PR 05-SEP-1997; 97US-0057650.



Db 601 TTTCCAGAAAGGACATCCCATCCAGCCAGAAACCCGCGAGATACAACTGGCCCTTCAG 660  
Qy 4362 AGATCGGGATGGTGGACAGATCTCCGACACAGCTGGGCTTCATCCGACAGACCA 4421  
Db 661 AGATCGGGATGGTGGACAGATCTCCGACACAGCTGGGCTTCATCCGACAGACCA 720  
Qy 4422 CATTGGGATCCCAACAGCAGCAGGCGCTGTGAACAACTCAGTGGCATAAACCGAAG 4481  
Db 721 CATGGGATCCCAACAGCAGCAGGCGCTGTGAACAACTCAGTGGCATAAAGAGAG 780  
Qy 4482 TCTGACCCGCGCTCGCCCTTATCACTCCCAAGGTTTCCACCGAGGAGTGAAGAT 4541  
Db 781 TCTGACCCGCGCTCGCCCTTATCACTCCCAAGGTTTCCACCGAGGAGTGAAGAT 840  
Qy 4542 GAACAAGTTCTGCTGTTGAGGACAGACTTTCTGGAAGCAGAGGAGCCACCTGAAA 4601  
Db 841 GAACAAGTTCTGCTGTTGAGGACAGACTTTCTGGAAGCAGAGGAGCCACCTGAAA 900  
Qy 4602 GGAGAGCAACAAGAGAGCTCTGAGCATTTGAGGCTTGGAACTCACAATCTGAGGAGGT 4661  
Db 901 GGAGAGCAACAAGAGAGCTCTGAGCATTTGAGGAGCTTGGAACTCACAATCTGAGGAGGT 960  
Qy 4662 GGACCACTTCCCTCTCTCCCTGCTTAAAGCAGCATGGGCGTCTTCTCCCTCTTC 4721  
Db 961 GGACCACTTCCCTCTCTCCCTGCTTAAAGCAGCATGGGCGTCTTCTCCCTCTTC 1020  
Qy 4722 CTTTCCCTTGCATGTGAATTAAGTGTGAAGAAATTCGCGTGGACATTTTCAGACTTGT 4781  
Db 1021 CTTTCCCTTGCATGTGAATTAAGTGTGAAGAAATTCGCGTGGACATTTTCAGACTTGT 1080  
Qy 4782 TGCCTTGAATGACAGTGCAGCATCTTCGAGCTCCAGCTGTGCTGCTGCCACATCAC 4841  
Db 1081 TGCCTTGAATGACAGTGCAGCATCTTCGAGCTCCAGCTGTGCTGCTGCCACATCAC 1140  
Qy 4842 ACAGTATCTTCCAAATTCACAAGATCATCACACAAGATGATCACTCTGCTGCTGCTTC 4901  
Db 1141 ACAGTATCTTCCAAATTCACAAGATCATCACACAAGATGATCACTCTGCTGCTGCTTC 1200  
Qy 4902 TCAATGCTGGAAGATTTTTTTTATCTCTCTTATATTTCAATTCACAGTCCAGCACT 4961  
Db 1201 TCAATGCTGGAAGATTTTTTTTATCTCTCTTATATTTCAATTCACAGTCCAGCACT 1260  
Qy 4962 TGAATCATTTGGGATTAAGAGAAAGCTAGCCATTTGAATCTACTTGGGCTTTAAACCCAC 5021  
Db 1261 TGAATCATTTGGGATTAAGAGAAAGCTAGCCATTTGAATCTACTTGGGCTTTAAACCCAC 1320  
Qy 5022 CAAGGAACAAGAAACAAATGAATCTTTGAGTACAGTCTTGTCCACTTGTTC 5081  
Db 1321 CAAGGAACAAGAAACAAATGAATCTTTGAGTACAGTCTTGTCCACTTGTTC 1380  
Qy 5082 AATGTCCTCTCTT -AAAAAATAAGATTTAAAGATTTGTTCAAGAGTAATAT 5139  
Db 1381 AATGTCCTCTCTTAAAAAATAAGATTTAAAGATTTGTTCAAGAGTAATAT 1440  
Qy 5140 AATTCATTTAATGATTAAGATTAATTTAAACCTTAAGTGGTGGCCACTGCTT 5199  
Db 1441 AATTCATTTAATGATTAAGATTAATTTAAACCTTAAGTGGTGGCCACTGCTT 1500  
Qy 5200 CTGAAACCAATATGCGGACAGGGTGTGGCCACACCAAGAAGCGGAGACCTGGC 5259  
Db 1501 CTGAAACCAATATGCGGACAGGGTGTGGCCACACCAAGAAGCGGAGACCTGGC 1560  
Qy 5260 TGTGACCCGCTGCTCCATGTCCTTCTGTCTCAACCCGGAAGTCCCTATCTGGAAG 5319  
Db 1561 TGTGACCCGCTGCTCCATGTCCTTCTGTCTCAACCCGGAAGTCCCTATCTGGAAG 1620  
Qy 5320 TATGAATGTTAGCAATTAATACAAAGACCTCATCTGCTCTTCCCAAGTGAAG 5379  
Db 1621 TATGAATGTTAGCAATTAATACAAAGACCTCATCTGCTCTTCCCAAGTGAAG 1680  
Qy 5380 GTTCTTCTGTAACGTTTGCATGCGCAGGGAGGGAACCTGCTGTGCTG 5439  
Db 5380 GTTCTTCTGTAACGTTTGCATGCGCAGGGAGGGAACCTGCTGTGCTG 5439

Db 1681 GTTCTTCTGTAACGTTTGCATGCGCAGGGAGGGAACCTGCTGTGCTG 1740  
Qy 5440 TGTGACCTTATGAGGACAGAGCTGTGATTTGGCGGATGTGTCCTCATTTAGATG 5499  
Db 1741 TGTGACCTTATGAGGACAGAGCTGTGATTTGGCGGATGTGTCCTCATTTAGATG 1800  
Qy 5500 GATGCAAAACCCATTTTAAATTAATTTCTTTGATTTTGTAAATTTAGAGGTAG 5559  
Db 1801 GATGCAAAACCCATTTTAAATTAATTTCTTTGATTTTGTAAATTTAGAGGTAG 1860  
Qy 5560 TTTTGTGTTTTG -TTTTGTTTTTTTTTAAAGAAACATTTAATGATTTAGATG 5618  
Db 1861 TTTTGTGTTTTGTTTTGTTTTTTTTTAAAGAAACATTTAATGATTTAGATG 1920  
Qy 5619 AGTGAAGAGCTGGGATGTTGAGCTAATGCCAGCTGTTAATAGCTCTTCAAGAC 5678  
Db 1921 AGTGAAGAGCTGGGATGTTGAGCTAATGCCAGCTGTTAATAGCTCTTCAAGAC 1980  
Qy 5679 AGCTTCCCTTTTATGAAATTTGCAATTTAGGAAATTAACAGCCTTTAACTGATTAAGAT 5738  
Db 1981 AGCTTCCCTTTTATGAAATTTGCAATTTAGGAAATTAACAGCCTTTAACTGATTAAGAT 2040  
Qy 5739 CAATAACCTGTTAGACATGCCAGCTTTGCAAGGCGAGTTAGTACCAAGACTAACC 5798  
Db 2041 CAATAACCTGTTAGACATGCCAGCTTTGCAAGGCGAGTTAGTACCAAGACTAACC 2100  
Qy 5799 CCAAGTGGCTTATGAGCCTGCATATAGAGAGCCTAAGTGTAGCAACATCTGTCA 5858  
Db 2101 CCAAGTGGCTTATGAGCCTGCATATAGAGAGCCTAAGTGTAGCAACATCTGTCA 2160  
Qy 5859 CAGCTGCTATTAACCTATATAGCTGAATGACATCCCTTCATCTATTTGTGTT 5918  
Db 2161 CAGCTGCTATTAACCTATATAGCTGAATGACATCCCTTCATCTATTTGTGTT 2220  
Qy 5919 TGCACAGACTCCGGAAGAGGAAGGCTGCCAATCTGAGTACTCAAAATGTAGAACT 5978  
Db 2221 TGCACAGACTCCGGAAGAGGAAGGCTGCCAATCTGAGTACTCAAAATGTAGAACT 2280  
Qy 5979 GCTGCTTGGATTTTTTTCATTAATTCAGCTGATCATTTGATGATGATTAAG 6038  
Db 2281 GCTGCTTGGATTTTTTTCATTAATTCAGCTGATCATTTGATGATGATTAAG 2340  
Qy 6039 TAAATACCTTCAATTTAAAGTGAATGAGCTTTTTCACAGTATCAACAAGT 6098  
Db 2341 TAAATACCTTCAATTTAAAGTGAATGAGCTTTTTCACAGTATCAACAAGT 2400  
Qy 6099 CAGTCTTATTAATTAATTCCTTCTGATCAATGCAATTTGTCTACTGCTTATTAAT 6158  
Db 2401 CAGTCTTATTAATTAATTCCTTCTGATCAATGCAATTTGTCTACTGCTTATTAAT 2460  
Qy 6159 TGTCAATTAATGATTTGATTTTACATGTAATGATGATTTGACAGTTTATTAAT 6218  
Db 2461 TGTCAATTAATGATTTGATTTTACATGTAATGATGATTTGACAGTTTATTAAT 2520  
Qy 6219 AGGCTATGAGCCTGATGCAATAGAAAGACAAATAGCTCTACCAAGTTGAC 6278  
Db 2521 AGGCTATGAGCCTGATGCAATAGAAAGACAAATAGCTCTACCAAGTTGAC 2580  
Qy 6279 AAATGTTATCAAGCAATTAAGTAATTTAGAACATAGAGCTGCTAATCTCAGTCTCT 6338  
Db 2581 AAATGTTATCAAGCAATTAAGTAATTTAGAACATAGAGCTGCTAATCTCAGTCTCT 2640  
Qy 6339 GTGATGCAAGTGCAGATGACAAATTAAGTGGATTTCTCATCTTTGATTAAGT 6398  
Db 2641 GTGATGCAAGTGCAGATGACAAATTAAGTGGATTTCTCATCTTTGATTAAGT 2700  
Qy 6399 TGTACCTGATGCTTTTGAAGACATTTGTTGAGAGTCTGATCCCTTTGATTTTAA 6458  
Db 2701 TGTACCTGATGCTTTTGAAGACATTTGTTGAGAGTCTGATCCCTTTGATTTTAA 2760  
Qy 6459 TACATAATTTGATATTTGTTATTTTGTGAAGATGTTAGAAATGATGATGTT 6518  
Db 2761 TACATAATTTGATATTTGTTATTTTGTGAAGATGTTAGAAATGATGATGTT 2820

OY 6519 ATGCTTACATCCAGTTGTACAGCTGGAAATATAATATATACAT 6568  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 2821 ATGCTTACATCCAGTTGTACAGCTGGAAATATAATATATACAT 2870

## RESULT 4

ABAB3333  
 ID ABAB3333 standard; cDNA; 2890 BP.

AC ABAB3333;

DT 07-FEB-2002 (first entry)

XX Human secreted protein gene 140 SEQ ID NO:150.

XX Human; secreted protein; immunomodulatory; antisclerotic; anti-HIV;  
 KW dermatological; immunosuppressive; antiinflammatory; immunostimulant;  
 KW cytosolic; cardiant; vascular; anti-angiogenic; ophthalmological;  
 KW neuroprotective; nootropic; anticonvulsant; antialzheimers; vulnerary;  
 KW antiparkinsonian; antimicrobial; gene therapy; vaccine; immune disorder;  
 KW multiple sclerosis; systemic lupus erythematosus; HIV infection; cancer;  
 KW human immunodeficiency virus; hyperproliferative disorder; wound healing;  
 KW Gaucher's disease; cardiovascular disease; Scimitar syndrome; chemotaxis;  
 KW Chaga's cardiomyopathy; coronary arteriosclerosis; angiogenic disorder;  
 KW corneal graft neovascularisation; diabetic retinopathy; regeneration;  
 KW neurological disorder; Huntington's chorea; Alzheimer's disease;  
 KW Parkinson's disease; infectious disease; chromosome 4; ss.

XX Homo sapiens.

OS MO200162891-A2.

XX 30-AUG-2001.

XX 21-FEB-2001; 2001WO-US05614.

XX 24-FEB-2000; 2000US-184836P.

XX 29-MAR-2000; 2000US-193170P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX NI J, Ebner R, Lafleur DM, Moore PA, Olsen HS, Rosen CA;

PI Ruben SM, Soppet DR, Young PE, Shi Y, Florence KA, Wei Y;

PI Florence C, Ha J, Li Y, Kyaw H, Fischer CL, Ferrite AM, Fan P;

PI Feng P, Endress GA, Dillon PJ, Carter KC, Brewer LA, Yu G;

PI Zeng Z, Greene JM;

PI MPI: 2001-625724/72.

XX P-PSDB; ABB50440.

XX Nucleic acids encoding 207 human secreted polypeptides, useful for

PT preventing, diagnosing and/or treating, e.g. cancers, Parkinson's

PT disease and diabetic retinopathy -

XX Claim 1: Page 1004-1005; 1533pp; English.

XX ABB50301 to ABB51287 and ABA83194 to ABA83441 represent human secreted  
 CC proteins (I) and polynucleotide (II) sequences. (I) and (II) have various  
 CC activities based on the tissues and cells the genes are expressed in.  
 CC Example of these activities include: immunomodulatory; antisclerotic;  
 CC dermatological; immunosuppressive; antiinflammatory; immunostimulant;  
 CC anti-HIV; cytosolic; cardiant; anti-angiogenic; ophthalmological;  
 CC neuroprotective; nootropic; anticonvulsant; antialzheimers; vascular;  
 CC antiparkinsonian; antimicrobial; gene therapy; vaccine; immune disorder;  
 CC in gene therapy and vaccine production. (I) and (II) can be used in the  
 CC prevention, diagnosis and treatment of immune disorders (e.g. multiple  
 CC sclerosis, systemic lupus erythematosus and human immunodeficiency virus  
 CC (HIV) infections), hyperproliferative disorders (e.g. cancers and  
 CC Gaucher's disease), cardiovascular diseases (e.g. Scimitar syndrome,  
 CC Chaga's cardiomyopathy and coronary arteriosclerosis), angiogenic  
 CC disorders (e.g. corneal graft neovascularisation and diabetic  
 CC retinopathy), neurological disorders (e.g. Huntington's chorea,

CC Alzheimer's disease and Parkinson's disease), infectious diseases and/or  
 CC for promoting wound healing, regeneration and/or chemotaxis. ABA83185 to  
 CC ABA83193 and ABB50300 represent sequences used in the exemplification of  
 CC the present invention.

XX SQ Sequence 2890 BP; 816 A; 631 C; 618 G; 821 T; 4 other;

Query Match 43.2%; Score 2840.2; DB 22; Length 2890;  
 Best Local Similarity 99.7%; Pred. No. 0;

Matches 2862; Conservative 3; Mismatches 2; Indels 3; Gaps 2;

OY 3702 TTATATGCTACAGCTCAGTAATTTCTTCACAGCAGACAGAACTTCCAGATCAG 3761  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 1 TTATATGCTACAGCTCAGTAATTTCTTCACAGCAGACAGAACTTCCAGATCAG 60  
 OY 3762 GGGGATCGCGCTCAGTTGATGCTGCTGACAGTGGCCGTGGAGCTGCATGCTCA 3821  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 61 GGGGATCGCGCTCAGTTGATGCTGCTGACAGTGGCCGTGGAGCTGCATGCTCA 120  
 OY 3822 AGTGGCTCCATATATATATACAGAGATCCAGCAGAGAAAGCTGGAGACTCTTCCA 3881  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 121 AGTGGCTCCATATATATATACAGAGATCCAGCAGAGAAAGCTGGAGACTCTTCCA 180  
 OY 3882 TTGGGCTACCTCAGTTGATATATTCAGGGATCTGAGTTATGAGCATCAAGCAGC 3941  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 181 TTGGGCTACCTCAGTTGATATATTCAGGGATCTGAGTTATGAGCATCAAGCAGC 240  
 OY 3942 CATATGAGCAAAATTTATGTTTCTGATCATAGCACAAGTATTAACAGGCAAAATCAAGT 4001  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 241 CATATGAGCAAAATTTATGTTTCTGATCATAGCACAAGTATTAACAGGCAAAATCAAGT 300  
 OY 4002 AGAGAGAGCTTTGAACAAGCCAGTCCCGAGCAAGCTGGCGTCTTCACAGTTACTGG 4061  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 301 AGAGAGAGCTTTGAACAAGCCAGTCCCGAGCAAGCTGGCGTCTTCACAGTTACTGG 360  
 OY 4062 GGAAGAGCTCAGAGAGTGCACAGGACATTAAGAGGAGGGGTGGAAAGATGTTTCC 4121  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 361 GGAAGAGCTCAGAGAGTGCACAGGACATTAAGAGGAGGGGTGGAAAGATGTTTCC 420  
 OY 4122 ATTGAAGCCGAAAGCAGTACCTTAAGCTGTGACTAGCAGAGAAACCAAGCTGTCTCCC 4181  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 421 ATTGAAGCCGAAAGCAGTACCTTAAGCTGTGACTAGCAGAGAAACCAAGCTGTCTCCC 480  
 OY 4182 ATGCTCTCCCAATAGTGTGGCATCAAGTACTACAAAGGGGCTCATTTGACAGAAAGAG 4241  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 481 ATGCTCTCCCAATAGTGTGGCATCAAGTACTACAAAGGGGCTCATTTGACAGAAAGAG 540  
 OY 4242 GGCAGGATGAGAGGCCCGCCGACCCCTCCGGGCTACATTTGGAATTTCCATTACTGAC 4301  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 541 GGCAGGATGAGAGGCCCGCCGACCCCTCCGGGCTACATTTGGAATTTCCATTACTGAC 600  
 OY 4302 TTTTCAGAAAGGCACTCCCATTCAGAGCCAGAAACCCCGGAGCTACAAAGTGGCCCTTCAG 4361  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 601 TTTTCAGAAAGGCACTCCCATTCAGAGCCAGAAACCCCGGAGCTACAAAGTGGCCCTTCAG 660  
 OY 4362 AGATCGGGATGTCGTCGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4421  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 661 AGATCGGGATGTCGTCGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 720  
 OY 4422 CATGGGATCCCAAGCAGCAGAGCCCTGTGAAACAACTCAGAGGCAATTAACAGGAGAG 4481  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 721 CATGGGATCCCAAGCAGCAGAGCCCTGTGAAACAACTCAGAGGCAATTAACAGGAGAG 780  
 OY 4482 TCTGACCCGCGCTCGCCCTTATCATGTCCTCAAGGAGTTTTCACAGGAGAGTGAAGAT 4541  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 781 TCTGACCCGCGCTCGCCCTTATCATGTCCTCAAGGAGTTTTCACAGGAGAGTGAAGAT 840  
 OY 4542 GAACAGTTTCTGCTGTTTGAAGCAGAGACTTTTGTGAAGCAGAGGAGGACCTGTAAG 4601  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 841 GAACAGTTTCTGCTGTTTGAAGCAGAGACTTTTGTGAAGCAGAGGAGGACCTGTAAG 900  
 OY 4602 GGAGAGCACAAGAGAGTCTGAGCTTGGAGCCTTGAAGTCAATTTGAGGAGAGGT 4661  
 ||||||||||||||||||||||||||||||||||||||||||||||||



Db 901 GGAGAGCACAAGAGAGCTCTGAGCATGGAGCCTTGGAACTCACAATTCCTGAGAGAGCT 960  
Qy 4662 GGAGCAGATTGGCGCTCTCTCCCTGCTTAAAGCAGCATGGGGCTCTCTCTCCCTTCTTC 4721  
Db 961 GGAGCAGATTGGCGCTCTCTCCCTGCTTAAAGCAGCATGGGGSTCTCTCTCTCCCTTCTTC 1020  
Qy 4722 CTTTCCCTTTGCACTGTGAATACTGTGAAGAAATTCCTGGCAGCTTTTCAACATTTGT 4781  
Db 1021 CTTTCCCTTTGCACTGTGAATACTGTGAAGAAATTCCTGGCAGCTTTTCAACATTTGT 1080  
Qy 4782 TGCCTTGAATGACAGTGCAGCATCTTGGAGCTCCCACTGTTGCTGCTGCAATCAG 4841  
Db 1081 TGCCTTGAATGACAGTGCAGCATCTTGGAGCTCCCACTGTTGCTGCTGCAATCAG 1140  
Qy 4842 ACAGTATCATTCCAAAATTCCAAGATCATCACACAGATGATTCACCTGCTGCAATC 4901  
Db 1141 ACAGTATCATTCCAAAATTCCAAGATCATCACACAGATGATTCACCTGCTGCAATC 1200  
Qy 4902 TCAATGCTGGAGAGATTTTAACTCTCTCTTGAATTCATTCAGCTGCTGCAATC 4961  
Db 1201 TCAATGCTGGAGAGATTTTAACTCTCTCTTGAATTCATTCAGCTGCTGCAATC 1260  
Qy 4962 TCAATGCTGGAGATTAATGAGAAAAGCTAGCATTTGAGGAGCTTTAAACCCAC 5021  
Db 1261 TCAATGCTGGAGATTAATGAGAAAAGCTAGCATTTGAGGAGCTTTAAACCCAC 1320  
Qy 5022 CAAGGAGACAAAGAAAACATGAAATCTTTGAGTACAGTGGCTGCTGCAATTTTAC 5081  
Db 1321 CAAGGAGACAAAGAAAACATGAAATCTTTGAGTACAGTGGCTGCTGCAATTTTAC 1380  
Qy 5082 AATGCTCTGCTTTT-AAAAAAAATGAGTTTAAAGATTTTGTTCAGAGATTAATAT 5139  
Db 1381 AATGCTCTGCTTTTAAAAAAAATGAGTTTAAAGATTTTGTTCAGAGATTAATAT 1440  
Qy 5140 AATGCTCTGCTTTTAAATGATTAATGATTTTAACTTAAAGAGGTTTCCAGAGCTGCTT 5199  
Db 1441 AATGCTCTGCTTTTAAATGATTAATGATTTTAACTTAAAGAGGTTTCCAGAGCTGCTT 1500  
Qy 5200 CTGAAAACCAAAATATGCGGAGCAGGAGTGGGACACCAAGAAAGAGAGAGAGCTGGC 5259  
Db 1501 CTGAAAACCAAAATATGCGGAGCAGGAGTGGGACACCAAGAAAGAGAGAGAGCTGGC 1560  
Qy 5260 TTGTGACCTGGCTTCCCATGCTCTCTGCTGCTCACCCGGAAGTGCCTATCTGGAAG 5319  
Db 1561 TTGTGACCTGGCTTCCCATGCTCTCTGCTGCTCACCCGGAAGTGCCTATCTGGAAG 1620  
Qy 5320 TATGAATGTTAGCCATTAATACCAAGACACCTCATCTGCTGCTTCCCAATGGATGGG 5379  
Db 1621 TATGAATGTTAGCCATTAATACCAAGACACCTCATCTGCTGCTTCCCAATGGATGGG 1680  
Qy 5380 GTTCTCTGTGAATACTGTTGACATGAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAG 5439  
Db 1681 GTTCTCTGTGAATACTGTTGACATGAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
Qy 5440 TCTGAGCCTTATGAGGAGCAGAGAGGCTGCTATGCGGATGCTGCTGCTGCTGAGAG 5499  
Db 1741 TCTGAGCCTTATGAGGAGCAGAGAGGCTGCTATGCGGATGCTGCTGCTGCTGAGAG 1800  
Qy 5500 GATGCGAAACCCCATTTTAAGTATATTTCTGATTTTGTAAATTTAGAGGTTAGG 5559  
Db 1801 GATGCGAAACCCCATTTTAAGTATATTTCTGATTTTGTAAATTTAGAGGTTAGG 1860  
Qy 5560 TTTTGTGTTTGTG-TTTTGTGTTTGTGTTTAAAGAAACATTTAACTGGATGAGATTCG 5618  
Db 1861 TTTTGTGTTTGTGTTTGTGTTTGTGTTTAAAGAAACATTTAACTGGATGAGATTCG 1920  
Qy 5619 AGTGAAGAGAGCTTGGAGATTTGGAGCTAATGCGAGCTGTTTACTGCTCTTCAAGAC 5678  
Db 1921 AGTGAAGAGAGCTTGGAGATTTGGAGCTAATGCGAGCTGTTTACTGCTCTTCAAGAC 1980  
Qy 5679 AGCCTCCCTTATTTGAATTTGGCATTTAGGAATTAACAAGAGCTTTAAACGATTAAGAT 5738  
Db 1981 AGCCTCCCTTATTTGAATTTGGCATTTAGGAATTAACAAGAGCTTTAAACGATTAAGAT 2040

Qy 5739 CAAAACCTGGTTAGACATGCCAGCCTTTGCAAGGACAGTTAGTACACAAAGACTAACCT 5798  
Db 2041 CAAAACCTGGTTAGACATGCCAGCCTTTGCAAGGACAGTTAGTACACAAAGACTAACCT 2100  
Qy 5799 CCAAGTGGCTTTATGAGAGCTGTCATATAGAGAGAGCTTAAGTGTAGACATCTGCTCA 5858  
Db 2101 CCAAGTGGCTTTATGAGAGCTGTCATATAGAGAGAGAGCTTAAGTGTAGACATCTGCTCA 2160  
Qy 5859 CAGCTGCTATTAACCTATTAATGAGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5918  
Db 2161 CAGCTGCTATTAACCTATTAATGAGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220  
Qy 5919 TGCACAGACTCCGGAAGAGTGAAGAGCTGCCAATCTGAGTACTCAAAATGTAGAACT 5978  
Db 2221 TGCACAGACTCCGGAAGAGTGAAGAGCTGCCAATCTGAGTACTCAAAATGTAGAACT 2280  
Qy 5979 GCTGCTCTGAGATTTTTCATTAATTCAGCTGATCATATTTGATCATGATGATTAACG 6038  
Db 2281 GCTGCTCTGAGATTTTTCATTAATTCAGCTGATCATATTTGATCATGATGATTAACG 6098  
Qy 6039 TAAATGCTTCAAAATTTTAAAGTGAATTCAGTGTGTTTCACTGATCAAAACATGT 6098  
Db 2341 TAAATGCTTCAAAATTTTAAAGTGAATTCAGTGTGTTTCACTGATCAAAACATGT 2400  
Qy 6099 CAGTGTATTTAATTAATTTCTCTCTGATCATGAGAGATTTGCTACATGCTTATTAAT 6158  
Db 2401 CAGTGTATTTAATTAATTTCTCTCTGATCATGAGAGATTTGCTACATGCTTATTAAT 2460  
Qy 6159 TGTCAATTAATGATTTGATTTTATTAATTAATTAATTAATTAATTAATTAATTAAT 6218  
Db 2461 TGTCAATTAATGATTTGATTTTATTAATTTTATTAATTAATTAATTAATTAATTAAT 2520  
Qy 6219 AGGCTATGACACATGATGATATAGAAAGACAGAAATCTAGCTTACACAAAGTTGCAC 6278  
Db 2521 AGGCTATGACACATGATGATATAGAAAGACAGAAATCTAGCTTACACAAAGTTGCAC 2580  
Qy 6279 AATGCTATTTAAGCATTAAGTAAATGTAAGACATGAGAGAGAGAGAGAGAGAGAG 6338  
Db 2581 AATGCTATTTAAGCATTAAGTAAATGTAAGACATGAGAGAGAGAGAGAGAGAGAGAG 2640  
Qy 6339 GTGATGTCAGAGTGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6398  
Db 2641 GTGATGTCAGAGTGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2700  
Qy 6399 TGTACCTGATGCTTTTAAAGAGACATTTGGTGAAGATGATGAGATGATGATGATGAT 6458  
Db 2701 TGTACCTGATGCTTTTAAAGAGACATTTGGTGAAGATGATGAGATGATGATGATGAT 2760  
Qy 6459 TACAATAATGATGATTAATTTGATTTTGTGAGATGATGAGAAATGATGATGATGAT 6518  
Db 2761 TACAATAATGATGATTAATTTGATTTTGTGAGATGATGAGAAATGATGATGATGATGAT 2820  
Qy 6519 ATGCTTTCATCCAGTTTGTACAGCTGGAAGTAATTAATTAATTAATTAATTAATTAAT 6568  
Db 2821 ATGCTTTCATCCAGTTTGTACAGCTGGAAGTAATTAATTAATTAATTAATTAATTAAT 2870

RESULT 5  
ABK94949  
ID ABK94949 standard; cDNA: 6383 BP.  
XX  
AC ABK94949;  
XX  
DT 30-AUG-2002 (first entry)  
XX  
DE Human novel polynucleotide #60.  
KW Human: gene; ss: inflammatory condition; shock; sepsis; immune response;  
cancer; wound healing; central nervous system disease; haematopoiesis;  
peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;  
myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;  
cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;  
KW



KM bone degenerative disorder; periodontal disease; reperfusion injury;  
KM lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;  
KM allergic condition; thrombolysis; thrombosis; coagulation disorder;  
KM fungal infection.  
OS Homo sapiens.  
PN MO200244340-A2.  
XX  
XX  
PD 06-JUN-2002.  
XX  
PF 30-NOV-2001; 2001WO-US47004.  
XX  
PR 30-NOV-2000; 2000US-0028952.  
XX  
XX (HYSE-) HYSEQ INC.  
PI Tang YF, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;  
PI Yamazaki V, Ujwal ML, Drmanac RT;  
XX  
DR MPI: 2002-508509/54.  
DR P-PSDB; AB666725.  
XX  
XX Novel nucleic acids and polypeptides for diagnosis, treatment of  
PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell  
PT disorders, cancer and promoting wound healing -  
XX  
PS Claim 1: Page 469-477; 672pp; English.  
XX  
XX The invention relates to human novel polynucleotides and associated  
CC polypeptides. The polynucleotides and polypeptides are useful for  
CC treating inflammatory conditions such as arthritis, nephritis, Crohn's  
CC disease, ischemia-reperfusion injury, shock, sepsis, immune responses  
CC and cancer and for promoting wound healing. The sequences are used to  
CC induce the proliferation of neural cells and regeneration of nerve and  
CC brain tissue, and are useful for the treatment of central and peripheral  
CC nervous system diseases and neuropathies, such as Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease and amyotrophic lateral  
CC sclerosis. The sequences are involved in chemotactic or chemokinetic  
CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid  
CC cell disorders and platelet disorders such as thrombocytopenia,  
CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of  
CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal  
CC disease. The sequences of the invention are also useful for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues, immune deficiencies and disorders  
CC including severe combined immunodeficiency (SCID), bacterial or fungal  
CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia  
CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis  
CC and coagulation disorders. Sequences ABK94890-ABK94982 represent human  
CC novel polynucleotides of the invention.  
XX  
SQ Sequence 6383 BP; 1953 A; 1236 C; 1435 G; 1759 T; 0 other:  
Query Match 24.2%; Score 1587; DB 24; Length 6383;  
Best Local Similarity 66.0%; Pred. No. 0;  
Matches 2559; Conservative 0; Mismatches 1205; Indels 99; Gaps 12;

DB 817 GGAAGCAGCAGTTTATCTGATATCTATTCAGGGCTACGGAGAGTGAGGATGATAGAT 876  
QY 285 CTGATGGGTTGGCACAACAGACGTGATTCGGAAGCAGCAGCATGANAAGCATT 344  
DB 877 TTGACACGCTTTCAGAAAGACCTGTTGATTCGAGCATATCACGAAGATGAAGAG 936  
QY 345 GAGAGAGCATTCAGATCCCTGATGAGCAGGAGCATTTGTGAGACCTCCAGAAAGGAC 404  
DB 937 ATTGATCGAACAAGATCCATTGCGAGGGCCAGATCTTGTTCGAAGATGCTTGAANAAG 986  
QY 405 CCAATTGACCGAGACAGATGATGATGAAACAATCTTGGAATTTATGACCAAGTTGCC 464  
DB 997 CCGCAGAGCAAAACCTGATGATGACATTTGAACAATTCCTGAGATTTGACCAAGCTCC 1056  
QY 465 GCTTTTCCCAATATGACATGTCATGAGGCGAAGAACTGTCTGCTGATGTTGGCA 524  
DB 1057 GCATTTGCAAAACATGACCATGCTGTGAGAGAGAACTGCTCAGTATGATTTTGAA 1116  
QY 525 GTGGTGAAGAGCAGGAGGACATAGTGTAAATGATGTGCAAGAGCTGCTGCTGCA 584  
DB 1117 GTGGTAGAGCAGGCTGAGCTATTTATTCCTGGAAGATGGCAAGAGCTTGACTATG 1176  
QY 585 GTGATTCCTCAATGATCTGTGAGAGTACTTATCCAGATGAAAGCAGAAATACTGTGC 644  
DB 1177 GTTATTTTAAAGCGCACTGTGGAATTCAGTCATCCAGATGAAAGTTGAATTTCT 1236  
QY 645 ATGGGAATATGTTTGGTGTCTCTCTACATGACACAAAGATATCATGAAGAGTATG 704  
DB 1237 ATGGAAATAGTTTGGAAATTTACTCCACTCTGATTAAGCATGATGAAATATGTC 1296  
QY 705 AGAACAAAGGTGATGATGCTGCGAGTTTGTCTGATAGCCAGCAAGATTCCTGCCAT 764  
DB 1297 AGACATTAAGTATGATGTTTGTGATGCTGATGCTGATGCTGATGCTGATGCTG 1356  
QY 765 CTCATCAATGATGAAAGAAAGAAATGGAAGAGAGAGAGATGTTATGCTG 824  
DB 1357 TTAACCATGTGGAAGAAATATACCATTAAGTTGAGGAAGAGAGAAATTTGATGTA 1416  
QY 825 AAGAACACCGAAGAACTGATGCACTGGAACAAGAAAGGACACATTTGATCAAGGT 884  
DB 1417 CATGACATCTGGGAAGCTAGACCGAGTGAACCAAGGAACACATTTGATCAAGGCA 1476  
QY 885 ACTCGAAGAAAGTTAAACAATGATTTGGTGAAGAGATCTCAGATGATTCACAAC 944  
DB 1477 ACACCTGAGCGCTCTCATATGATTTAATGAAGAAATTCATCGTGATCCAACTTAT 1536  
QY 945 ATGAAGACTTCTGTGACCTATGACCTTCTTCTTGAAGGATGAGAGTGGGCAAA 1004  
DB 1537 ATGAAGATTTTCTATTAATTAAGTACAGACATTTCTTGAAGTCTTGTGATGTTGG 1596  
QY 1005 AACTTATGAGTGTGTTAATGACCCGAGCTCAGGATATAGTTTACAGGGTATATTA 1064  
DB 1597 AACTTATGGAATGTTTAAAGATGACAGCTTAAGATTAAGTATGACAGCATTTGATTA 1656  
QY 1065 TTGTTGGTAAATATATCACTTCAATGATTTGAAGAGATCTCCGAAATGATCCATTTTA 1124  
DB 1657 TTATGGTAAATATATCACTTAAATGATTTGAAGGTACCTCTGATGATCCATTTCTTA 1716  
QY 1125 GAAGAATTTGAAAAACATCTGGAAGAGAGAAATGAGTGGAGACCTTAAGGCTGTTGA 1184  
DB 1717 GAGAGATTTGAAAAAATCTGGAAGATATCAAAAGATGAATGTCATCGCGTTATTGAAT 1776  
QY 1185 ATGCGGTGTCTCTTAAGCAAAAAGAGATTTGATGACGTTAACAAAACATCCCGAGA 1244  
DB 1777 ATTGCGTGTGCTCAAAAGGCTAAGTGAAGACAGTGTGTCTGCAAAAAGCTCCCGAG 1836  
QY 1245 GCTCCTTTGCTTTTCTTACTTGTGAGAGCTTGAGAGAGATTTGGAATCTTTGTGAC 1304  
DB 1837 TCCCTCTCAATATTCAGCTTAATGAGAGAGTGAAGAGGATTTGTTATTTTGTGAA 1896  
QY 1305 AGTGTAGATTCAGTAGCAAGCAACTGAAGCAGGCTTGGAAGCGGGGGATCAGATATTA 1364  
DB 1897 GAGTGAAGACCTGTTAGCAAAAGCTCGATTTACAGACTGAAGCTGATGAATGATATG 1556

Qy	1365	GAAAGTAAATGGCCAAACACTTTGAAAACATTCAGCTGTGCTAAAAGGCTATGGAATTTCTTGA	1424
Db	1957	GAAGTAAATGGCAAAACCTTTGGAGAAATTATCATTTATGAAAGCCGTTGAAATTTTGGAGG	2016
Qy	1425	AATAACACACATTTATTCATCAGTCGTGAAGAACCAATTTATTTGATTTAAAGAACCTTCTA	1484
Db	2017	AATTAATACCATCTTGCACTTACTGTGAAGACCAACATTTTGTGTCAAAAGATTTACTT	2076
Qy	1485	ACAAAGTTGTCAAGAGAGAAAAGAAATGCTGCCGCCACCTTCTTAAATTTGGTGACATT	1544
Db	2077	TTTAG --- GACTGAAACAAAGAAATCTGGTGTCTCCATATTTCCCAAAATTTCTGAAAAA	2133
Qy	1545	AAAAAGCCAGTCGCTACCTACCTCCAGATCTTTGGTGATGTGAAGACGTGATAGGA	1604
Db	2134	AAAAAGT --- AATCGCATTTCTATCCAGCATGTGCCAGGAGATATTGACAA --- GACATCA	2187
Qy	1605	CTTGAAAAAGTGAACAAAAAGTAAAGCCAACACTGTG --- GGAGGAAGAACAAAGCTG	1661
Db	2188	CAGAGAAAGGAAGTAAAGAAAGTTAAGCAAAATACTGTTTCAGGTGGAAGAAACAAAAATC	2247
Qy	1662	AAAAAGATCTGCACAAGACTGGATCACTATCTTGTCCACAGAAACCATACATGATATT	1721
Db	2248	AGGAATTTTGGATTAACACACATTAGTATCTTCCCTCCAAAGGTATTTAGTGATAGA	2307
Qy	1722	GGGATTTGGTCAAGTCGACAGTGCACATAGTATGATTAAGGACAGCAAAAGCACATCCCA	1781
Db	2308	GGGCTTAAGCCAAATCACAGATGACACATTTGTGGAAACAAAGGACATGTAGCATGTG	2367
Qy	1782	ACTGCATTTGCCCTGTCAAGTGGAACCTTATCATCCAGTATATCTGATTTATTGCAGTACAT	1841
Db	2368	GCTATATAGCCCATCCCTGGAAACACTCTCTCCAGACAGCCCTATCTCCGACGCTAC	2427
Qy	1842	CATCCGATTTTAACTTCACTCAAGTGTACTCCGACTCCGACATCAAGGCTAAAGGTTTGT	1901
Db	2428	ACCAAGTATGTTGGATTTTCCAAATCCTTCAGATATCCCTGATCAAGTATTAAGAGTTTTC	2487
Qy	1902	AAGGCTGATCAGCAAAAGCCGCTACATCATATGATAGTAAGACACTACGCAAAAGAGTG	1961
Db	2488	AAAGTGTATCAGCAAAAGTTGCTCATATATCATAGTAAAGACACACAGCTAAAGAAAGTA	2547
Qy	1962	GTCATTTAGGCTATCAGAGGAGTTTGCCTGATACGCCACCCGGATCAATTTTCACTATGT	2021
Db	2548	GTTTTATATGCTGTTCAATGAATTTGGTTTTCACCGGTGCATCCGACACATATTTCTCTGT	2607
Qy	2022	GAGGCTCTGTGCACACCTGAGGGAGTATCAAAACAAAGAAAGCTTCCAGATCAGCTTTC	2081
Db	2608	GAATTTCTGTACTCTCGTAGGGGTGTCAATAAACAGAAAGACTTCCAGATCAATCTTCC	2667
Qy	2082	AAACTTCAGACAGATATCAACTGATGTGAAGGTATATTCTGAAAAAACACATGGAAGAA	2141
Db	2668	AAATTTAGCTGTATGAATTCACATCACTCAATGGAAGGATTTACTTTAAAAAATATACATGGAACAA	2727
Qy	2142	GAATCTCTTTTGTGATGATGAAGATGCTCAGAGATTTGTTGAGAGAGAGTCAAAATTTCCCTC	2201
Db	2728	GAATCCTTATTTTCAGATGAAGATGCTCAGAACTAGTTAAGGAAGGACACTATTCATG	2787
Qy	2202	CTTCAGCTCAGACACTGTGGAAGTTGCAACACAGCTCTATGCGAAATTTTGAACCTTT	2261
Db	2788	CTGCAGCTCATACTCATGAGGTGGCCACCAGCTGTCTCAATGAGGAGATTTGATTTGTTT	2847
Qy	2262	CGCAACATTTGAACCTGATGATATATATGATGATTTATTAACTCAGATCAAAAACAGC	2321
Db	2848	CGTATATATTGAACCGACAGTACATGCTATACCTTTTAAAGTTAAATTTCCAAAAACAGA	2907
Qy	2332	TGTGCCAATCTGAAAGAGATTTTGAAGAAATCTTAACACGAAGAACTTTTGGTAAACATCT	2381
Db	2908	AATACTCATTTTGAAGAGGTTTGAGACATTTGAACCAAGAGACATCTCGGGTGGCTCA	2967
Qy	2382	GAATTTCTCAGAGAAACAACACAGCTGTAAGAGATGAAGATCATTAACCATTTTCAATCAAG	2441
Db	2968	GAATTTTAACTGAAGCAATACGCTCAACGAATGAAGAGATTTATTATAGCATTTTATTAAA	3022

OY	2442	ATAGCACTGCACGTGTAAGGAATGCAGAATAATTTAACTCATGTGTTGCATCATCATCAGTGGC	2501
Db	3028	ATTGCACCTTCATTGCGAAGATGTAAGAACTTCAATTCATGTTTGCAATTAATAAGTGGC	3087
OY	2502	CTAAACCTGGCACCACTGGCAAGACTGTGGAACGACCTGGSGAAGAACTCCCAATTAATAC	2561
Db	3088	TTGAACTGGGCATGTGTGCAGAGACTCAGAGAGACTTGGGAAAAGTTGCCAAGCAAATAC	3147
OY	2562	GAAAACTATTTCAAGATCTCCAGACCTGTTGATCTTCCAGAAACATGGCAAAATAT	2621
Db	3148	GAGAAACATCTTCAGATCTACAAACAATTTTGTATCATCTAGAAACATGGCAAGATAT	3207
OY	2622	CCTAATGTTCTAATATAGTCAAAATCTACAACTCCCATTAATCCCTATTTCCCACTATTC	2681
Db	3208	AGAAATATTTCTTAGTAGTCAAGATATGACGCTCCCAATTTATTCACACTCTTCCCTGTGTC	3267
OY	2682	AAAAAGATCTCACCTTCCCTTCAGAGAAATGACTCAAAAGTAGACGGGTGTCAAT	2741
Db	3268	AAGAAAGATATGACATTTCTACATGAAAGAAATGACTCCAAGTAGATGTTTACTTAAC	3327
OY	2742	TTTTGAGAGCTTAAGATGATTTGC AAAAGAAATTTCTCACGTTGGCCGAAATGGCTTCAGTG	2801
Db	3328	TTTGAGAACTTAAGATGATTTGCCAAGAAATCCGCAAGTTGTTCGAAATGCTTGTCT	3387
OY	2802	AACATGGACCCCTCCCTCATGTTTCAGAGACTCGGAAGAAATAATGGCGAGTTTGGGGTCT	2861
Db	3388	AACATGGACCCCACTATGATGTT-----TCGACAGAGGTCA	3423
OY	2862	CTCAGCCAGGGTAGTACAAATGCACAGACATGTAGATGTTGTCACAGAGTGTCTATAA	2921
Db	3424	CTGAGCTCAGAGAGACCAAAATTCMAACATGTGTGATG---TCAGGAGGTGTCTCAANA	3480
OY	2922	AACGGGTACGTGTGTGTTCTTTCATGANTGCCAAAAGCTTATTAAGATGCCCAAAAG	2981
Db	3481	AAAAAGGCACGGCGCAGCTCTTGCTTTAATGCCAAGAGCTAATAGAGTGGCCAAATG	3540
OY	2982	GCTCGAAAGTGAAGCAGTACCTTTCCAAATTTGGAGCTAGAAATGACAGAGAGACTTT	3041
Db	3541	GCAAGGAAGGTGAAGCAGTATCTTTCCAGTCTGATGTAGACACAAATGGAGAAAGTTC	3600
OY	3042	CAGACATTAATCTCGACGTGTGAGCCGACCAACACATTTGGCTAAGATCCGTGTGAC	3101
Db	3601	CAGATGATGTCAATTACAGTGGGACCTGCAATGTGTAATCTTGACCAAGAAATTTAAGTGA	3660
OY	3102	AAAAAGCCTGTCAA--ATCCGAGACCTTCCAGTAGCTCCAAAGGCGAGGCTACAACAG	3158
Db	3661	AAAAGATCAGCAAAATCATCTGAATATGTCACAGTGCCTATGAGCTCAGCTGGCCAAACA	3720
OY	3159	AAAGCTAGTCCCCTGCCACAGCCCCACAGAGAGCCACCAGACCTTAATAATCAACAG	3218
Db	3721	ACTAAAGCCCACTTGCAATCAACCC-----CACAGAATAGAGCAG	3759
OY	3219	GGACTACAGGTTCGCCGCGGTGTCCTTTATTCCTTACAGGAGAAAGTGCCCGTAAGGAT	3278
Db	3760	GTGCTTCAGGTGCAGCGTGAATTTTGACCCCCATCAGAAAGGAGCAACAAAAGAC	3819
OY	3279	CTCCACCTTTTGGCATTAACCTCTGCCACAGACTTTAAAAAAAATTTCTTCTGTGTGAA	3338
Db	3820	CCTGACCTGAATACAAATTTACCTCAGAAATTTTITAGAACAA-----CTGAA	3867
OY	3339	GAGGAATTTGGAAGCTCACAGCAACAAAGCGCTCAAGATATCAATATCAAAATGACTTTCG	3398
Db	3868	GAAATTA-----GTGTGAAGAGATATCAGAAAGCACTAATTTCTGTGGCTCATCT	3918
OY	3399	CAGCTTTCTCTCCTCACTTCTTCACAGAGTTCTCCAAAGAAAGGCTATACTTTGGCT	3458
Db	3919	TTACTTTCTACTCGCTCGCATCTCTCCMAAGCTCCCTCCAAAGGTTTACACTTAAT	3978
OY	3459	CCCACTGCTACTGTGAAATTTTACATTTCTGGTCACTGAAATTTCTTCACGATTC	3518
Db	3979	CCATCAGCTAAATCTGACAACTGTGTCTAGCTCAGCCATATGTGAATTTCTTCACGGTCC	4038
OY	3519	AGTATGTTAGCAATTCGCTTTTGACTCAGTCCGACGTCTCACTGCACGATGAGAGGCGC	3578



705 AGAACAAGGTGATGCTGACCTGATTTGCTGCTAGTACCCAGCAGATTTACTGCCGATTT 764  
1297 AGGCTAAGTACATGATGATTTGCTGCTGATAGCCAGCAGATTTATGGAGAAATTT 1356  
765 CTCATCAAGTAGAAAGAACATGCAAAAAGTTGAAGAGAGAAAGAAAGTTGTTAGGGT 824  
1357 TTAAACCATGTGAAAAAATACCCTAAAGTTGAGAAAGAGAAAGAAATGTTATGTA 1416  
825 AAAGAACGAGAGAACTTGATCGAACTGAAACAAGAGGACATTTGTCATCAAGGGT 884  
1417 CATAGCACTGCGAGAACTAGACCGGAGTGAACCAAGAAAGACATTTGATCAAGGCA 1476  
885 ACCTCAGAAAGGTTAACAATGATTTTGTGGAAGAGCATTCATAGATTCACATTC 944  
1477 ACACCTGAGCTCTCATATGTCATTTAATAGAAAGAACATTCATCGTGATCCAACTTAT 1536  
945 ATAGAAAGCTTTCTGTTGACCTATAGACCTTTTCTTCTAGCCCAATGGAAGTGGGCAAA 1004  
1537 ATAGAAAGATTTTCTATTATTAACCTTACAGGACATTTCTTGAAGCTCTTGGATGGGATC 1596  
1005 AAGTTATTTGAGGTGTTTATGACCCGCTCAGGATTAAGTTACACGGGTAGTATTA 1064  
1597 AACCTAATGGAATGTTAAGTACAGCTTAAGAGTAAAGTGAACGAGATTTGATTA 1656  
1065 TTGTGGTAAATTAATCACTTCAATGATCTTTGAGAGATCTCTGCAATGATCGATTTTAA 1124  
1657 TTATGGGTAATTAATCACTTATTAATGATTTTGAAGTGAACCTGCTGATGATCGATTTCTA 1716  
1125 GAAGAATTTGAAGAAACATCTGGAAGAGAAAGAAATGAGTGAACCTTAAGCTTTGAAT 1184  
1717 GAGGAATTTGAAGAAATCTGGAAGATACAAAGATGAATGATCTCGGTTATTTGAT 1776  
1185 ATGCGCTGCTGCTTAAGCAAAAGAGATTTGATGAGTTTACAAACCATCCGAGAA 1244  
1777 ATGCGCTGCTGCTTAAGCAAAAGAGATTTGATGAGTTTACAAACCATCCGAGAA 1836  
1245 GCTCCTTTGCTCTTATCTTACTTGAGGCTCTGAGAGGATTTTGAATCTTTGTATAC 1304  
1837 TCCCTCTCAATTTAGCCCTTAATGAGGAGATGAGAAAGGATTTGATTTTGTGAA 1896  
1305 AGCTGATTTCAAGTGAAGCAAGCACTGAAGCAAGCTTGAAGCGGAGGATCGATTTAA 1364  
1897 GAGGTGAAGCTGTTGAGCAAGCTGCTGATTCAGAGCACTGAAGCGGATCGATTTAA 1956  
1365 GAAGTAATGCGCAAAACCTTTGAAACATTCAGCTGCAAAAGCTATGGAATTTCTTGA 1424  
1957 GAGTAATGCGCAAAACCTTTGAGAAATTTACATTTATGAAAGCCGTTGAAATTTTGAG 2016  
1425 AATTAACACATTTATCTATCACTGTAAGAAACCAATTTTGTATTTAAAGCACTTCTA 1484  
2017 AATTAATCACTATCTGACTTACTGATGAGACCAACATTTTGTCTTCAAAAGTACTT 2076  
1485 ACAGATTTGTCAGAGAGAAAGAAATGAGTGGCCCACTTCTTAATTTGCTGACATTT 1544  
2077 TTTTGG---GACTGAGCAAGAGAAATCTGATCTCTCATTTTCCAAAATTTGCTGAAAA 2133  
1545 AAAAAGGCCAATGCTACTCCATTTCCAGATCTTCTGATTAATGTAAGAAAGGATGATAG 1604  
2134 AAAAG---TAAATGCCCATTTCTATCCAGATGTCGAGAGATTTTGACA---GACATCA 2187  
1605 CTTGAAAAAGTGAACAAAAAGTAAAGCAACACTGTG---GAGAGAGAGAAAGAGT 1661  
2188 CAGGAGAAAGAGAGTAAAGAAAGTTAAAGCAATTTCTTTCAGGTGGAAGAAACAAATG 2247  
1662 AAAAAGATTAATCGACAGACTGATCGATGATCTTGGCAGAGAAACCATATCAATGATAT 1721  
2248 AGGAAGATTTTGGATTAAGAAACAGATTTAGTATCTTGGCTCCAAAGCTATTTAGTGGGA 2307  
1722 GGGATGTTGCTCACTGATGATGACAGCATAGTAAAGGAGAGAAAGCAATCCCA 1781  
2308 GGCCTAAGCCATTCACAGATGACAGCTTTGTGGAGAACAGGACATGATAGCATATGCTG 2367  
1782 ACTGATTTGCTGCTCACTGATGAGAACTTATCATCCATTAATCTGATTTATGCACTACAT 1841  
2368 GCTATTAATGCCCATCCCTGGAGAACTCTCAATCCAGAGAGCCCTGATCTCTGCAAGCTTAC 2427  
1842 CATGCAATTTTAACTTCACTGATGCTGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1901  
2428 ACCAGATTTGATTTTTCATTAATCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 2487  
1902 AAGGCTATCGAAGAGCCCTATCATGATGATGATGATGATGATGATGATGATGATGATGATG 1961  
2488 AAAGGATTTGAGCAAAAGTGTATCATATATCATATGATGATGATGATGATGATGATGATGATG 2547  
1962 GTCATTCAGCTATCAGGAGATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2021  
2548 GTTTTCAATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2607  
2022 GAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2081  
2608 GAAATTTCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2667  
2082 AAACCTGCAAGCAATATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2141  
2668 AAATTAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2727  
2142 GAAACTCTTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2201  
2728 GAAACTCTTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2787  
2202 CTTGACCTGACAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2261  
2788 CTTGACCTGACAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2847  
2262 CGCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2321  
2848 CTTGATTTTCAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2907  
2322 TGTGCAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2381  
2908 AATTAATCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2967  
2382 GAAATTTCTGAGAAACAAACAGCTGAGAGAGATGATGATGATGATGATGATGATGATGATG 2441  
2968 GAAATTTTAACTGAGAAATCAAGCTTCAACCAAGATTAAGATTTTAAATTTTAA 3027  
2442 ATAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2501  
3028 ATTCACCTTATGCTGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3087  
2502 CTAACCTGACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2561  
3088 TTGAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3147  
2562 GAAAACTATTTCAAGATTTCTCAAGACCTGTTGATCTTCCAGAAACATTTGCAATTTAT 2621  
3148 GAGAAACATTTCTCAAGATTTCTCAAGACCTTGTGATCTTCAATTTGCAATTTGCAATTTG 3207  
2622 CTTATTTCTCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2681  
3208 AGAATTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3267  
2682 AAAAAGATTTCACTTCTTCAAGAGAAATGATGATGATGATGATGATGATGATGATGATGATG 2741  
3268 AAGAAAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3327  
2742 TTTGAGAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2801  
3328 TTTGAGAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3387  
2802 AACATGAGACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2861  
3388 AACATGAGACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3423  
2862 CTCAGCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2921

```

Db 3424 CTGAGTCAAGGAAGCAACAATTCAAACATGCTGATGT---TCAGGAGAGTGCTCAACAA 3480
QY 2922 AACGGGTACGTCGTGTGTTCTTTCATTCGCAAAAAGCTTTATGACATGCCCAATG 2981
Db 3481 AAAAGGGACGGCGCGAGCTCTGCTTAATGCGCAGAAAGTATATGAGTGCCTCAAAATG 3540
QY 2982 GCTGSAAGAGTGAAGCAGTACCTTTCATTTTGGAGTAGAATGAGCGAGAGTCTT 3041
Db 3541 GCAGAGAGAGTGAAGCAGTATCTTTCCAGTCTGATGTAGAGCAGATGAGGAGAACTTC 3600
QY 3042 CAGACATTAATCTCTGAGTGTGAGCCAGCAACACATTTGCTTAAGATCCTGCTGAC 3101
Db 3601 CAGATGATGATCATATACAGTGGAGAGCTGCATATGTATGATACCTTGACCAAGATTTAAGTGAG 3660
QY 3102 AAAAACCCTGTCA-----ATCCGAGACCTCTCCAGTACTCCAAAGGCGAGGTCACAA 3155
Db 3661 AAAAGATCAGCCCAAGNNATCATCTGAATGTCTCAGTGTGATGAGTCACTGAGCCAA 3720
QY 3156 CAGAAAGCTCAGTCCCTGCCACAGCCCGACAGCAGCAGCACCACACATTAATCAAC 3215
Db 3721 ACAACTAAAGCCCATTTGCATCAACC-----CACAGATGAGC 3759
QY 3216 CAGGAGCTACAGGTTCCCGCCGTGCTCTTATCTTCAGGAGAAAGTGCCTGAAG 3275
Db 3760 CAGGTGCTTCAGTGGCAGAGCTGTTAATTTGACCCCATCGAAGAAAGGACAAACAAA 3819
QY 3276 GATCTCCACCTTTTGGCATAACTCTCCACAGCTTTAAAAAAATCTCTTGTGCT 3335
Db 3820 GACCTGCACTGATACAACTTTCCTCAGAAAGTTTGAAGCAAA-----CT 3867
QY 3336 GAGAGAGAAAGTTTGAAGCTCAGAAACAGGCTGAAGATCAATATCAATATGATCT 3395
Db 3868 GAAGAAATTA-----GTGTAAGAACATACAGAAACACTATTTGTGGGCTCA 3318
QY 3396 TCCAGCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3455
Db 3919 TCTTTCATCTGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3978
QY 3456 GCTCCAGTGTGCTGTGATTAATTTTTCAGATTTCTGCTACAGTGAATTTCTTCACGA 3515
Db 3979 ATTCCATTCACACTAAATCTGACAACTGTCTGACTCCAGCCATGTGATTTCTTCACGG 4038
QY 3516 TCCAGTATTTGATGACATTTGCTTTTGACTCACTGCTGCTGCTGCTGCTGCTGCTGCTG 3575
Db 4039 TCCAGCATCTGAGCAATTTCTCTGTTGACTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4098
QY 3576 CGCCAGAGGCAATCTCTGACATCTGGAACAAACCTAGGATGGGACAGATGAGAGG 3635
Db 4099 TGTTCCTCTCAGGCGCTGGAGATCCCGAATCCAGTGGGCAATTTGGAAGACAGAGCAC 4158
QY 3636 CGACCATGATTTAACCCTGATAGTATAGCTTGGGCTCTATGACCAATGTCCGAGGGC 3695
Db 4159 GCTTCAGGAGTAGAGATCTATGATCAACATGAGCCCTGGGTGAGACACTCTTCCAGCCATCT 4218
QY 3696 CGAGGCTTATATGCTACAGCTACAGTAAATTTCTTCCAGCAAGAGGAACTTTCCAG 3755
Db 4219 CTATATCAAGTGTATAGTGTCTCA-----TCGTCTGTGAGCATATGAGGATTTCTCAA 4272
QY 3756 GATCAGGAGGAGTGGCGGTCACTTGTATGCTGTGACAGTGGCGGTGGAGCTGAGCTCA 3815
Db 4273 GAGCATATCATTT-----ATAGAAGCAGTGAAGTGTGTGTGTAATTTGACTTCG 4323
QY 3816 TGTCTCAAGTGTCTCCATGATTAATATACAGACATCCAGCAGCAGAGAGAGCTGGGA 3871
Db 4324 TGTTCAGCAGCTCCCATGTGACAACTTCCAAAGCTTCCAAAGCCAAAAGCTGGGA 4379

```

RESULT 7  
 AAL20515  
 ID AAL20515 standard; cdna; 481 BP.  
 AC AAL20515;

```

XX 07-DEC-2001 (first entry)
DT Human breast cancer expressed polynucleotide 12972.
DE Human; breast cancer; cell marker; cytosolic; ss.
KW Homo sapiens.
OS WO200151628-A2.
PN 19-JUL-2001.
PD 10-JAN-2001; 2001MO-US00798.
PE 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
PX (MIL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA Lillie J, Xu Y, Wang Y, Steinmann K.
PI WPI; 2001-451856/48.
PT New peptide useful as a marker for the diagnosis of breast cancer
PS Claim 1; Page 2298; 3695pp; English.
XX The invention relates to human breast cancer expressed polynucleotides
XX (AAL07544-AAL26789) and methods of assessing whether a patient is
XX afflicted with breast cancer by examining the correlation between the
XX expression of certain markers and the cancerous state of breast cells.
XX The polynucleotides and encoded polypeptides are potential markers for
XX detecting, diagnosing, monitoring, characterizing treating and
XX potentially preventing breast cancer. The polynucleotides and encoded
XX polypeptides are also useful for isolating compounds with cytostatic
XX activity.
XX Sequence 481 BP; 141 A; 136 C; 96 G; 108 T; 0 other;
SQ
Query Match 7.2%; Score 471; DB 22; Length 481;
Best Local Similarity 100.0%; Pred. No. 3.1e-108;
Matches 471; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2999 GTACCTTTCGAATTTGGAGCTAGAAATGAGAGAGAGTCTTACGATTAATCTGCA 3058
Db 9 GTACCTTTCGAATTTGGAGCTAGAAATGAGAGAGAGTCTTACGATTAATCTGCA 68
QY 3059 GTGTGAGCCAGCAACCAACCATTTGCTTAAGAAATCTGTGACAAAGCTGCAATC 3118
Db 69 GTGTGAGCCAGCAACCAACCATTTGCTTAAGAAATCTGTGACAAAGCTGCAATC 128
QY 3119 CGAGACTCTTCAGTACCTCAAGGCGAGGCTCAACAGAAAGCTGCTCCGCCACA 3178
Db 129 CGAGACTCTTCAGTACCTCAAGGCGAGGCTCAACAGAAAGCTGCTCCGCCACA 188
QY 3179 GCCCGAGCAGCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3238
Db 189 GCCCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 248
QY 3239 GTCCTTTATCTTTCAGAGAAAGTCCGTAAGATCTCCACCTTTTGGCATAAA 3298
Db 249 GTCCTTTATCTTTCAGAGAAAGTCCGTAAGATCTCCACCTTTTGGCATAAA 308
QY 3299 CTCCTCAGCAAGCTTTAAAAAAATTTCTTCTTGTCTGAAGAAGAGAGTTGGAACGTCA 3358
Db 309 CTCCTCAGCAAGCTTTAAAAAAATTTCTTCTTGTCTGAAGAAGAGAGTTGGAACGTCA 368

```

```

QY 3359 CAGAAACAGCTGAGATACATATCAATGATCTTCGACGCTTCTTCCTCTAC 3418
      |||||||
Db 369 CAGAAACAGCTGAGATACATATCAATGATCTTCGACGCTTCTTCCTCTAC 428
      |||||||
QY 3419 TTCTCCACAGAGTTCTCCAGAAAGGCTATCTTGGCTCCAGGCTAC 3469
      |||||||
Db 429 TTCTCCACAGAGTTCTCCAGAAAGGCTATCTTGGCTCCAGGCTAC 479
      |||||||

RESULT 8
AAL1617
ID AAL1617 standard; cDNA: 467 BP.
XX
AC AAL1617;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 4074.
XX
KM Human: breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PE 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192089.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steilmann K;
XX
DR WPI: 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer
PS Claim 1: Page 744; 3695pp; English.
XX
XX The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 467 BP; 139 A; 131 C; 93 G; 104 T; 0 other;
Query Match 7.0%; Score 459; DB 22; Length 467;
Best Local Similarity 100.0%; Pred. No. 3.3e-105; Indels 0; Gaps 0;
Matches 459; Conservative 0; Mismatches 0;

QY 2999 GTACCTTCCAAATTTGAGCTAGATAATGACGAGAGCTTTCAGACATTTATCTCTGCA 3058
      |||||||
Db 9 GTACCTTTCCAATTTGAGCTAGATAATGACGAGAGCTTTCAGACATTTATCTCTGCA 68
      |||||||
QY 3059 GTGTGAGCAGCAACACACATTTGCTTGAATCTGTGACAAAAGCCTGTCAATTC 3118
      |||||||
Db 69 GTGTGAGCAGCAACACACATTTGCTTGAATCTGTGACAAAAGCCTGTCAATTC 128
      |||||||
QY 3119 CGAGACCTCTCCAGTACTCAAGGCGAGGTACAAACAAAGCTCAGTCCCTGCCACA 3178
      |||||||

```

```

Db 129 CGAGACCTCTCCAGTACTCAAGGCGAGGTACAAACAGAAAGCTCAGTCCCTGCCACA 188
      |||||||
QY 3179 GCCCAGACAGAGCCACCACACATATAATACACGAGGACTACAGTTCGCCGCT 3238
      |||||||
Db 189 GCCCAGACAGAGCCACCACACATATAATACACGAGGACTACAGTTCGCCGCT 248
      |||||||
QY 3239 GTCCCTTATCCCTTCACGAGAAAGTCCCGTAAGGATCTCCACCTTTGGCATAAA 3298
      |||||||
Db 249 GTCCCTTATCCCTTCACGAGAAAGTCCCGTAAGGATCTCCACCTTTGGCATAAA 308
      |||||||
QY 3299 CTCTCCACAGGCTTTAAAAAATTTCTTCTTGTCTGACAGAGAGATTTGGAACGTCA 3358
      |||||||
Db 309 CTCTCCACAGGCTTTAAAAAATTTCTTCTTGTCTGACAGAGAGATTTGGAACGTCA 368
      |||||||
QY 3359 CAGAAACAGGCTGGAAGATACATATCAATGATCTTCGACGCTTCTTCCTCTAC 3418
      |||||||
Db 369 CAGAAACAGGCTGGAAGATACATATCAATGATCTTCGACGCTTCTTCCTCTAC 428
      |||||||
QY 3419 TTCTCCACAGAGTTCTCCAGAAAGGCTATCTTGGC 3457
      |||||||
Db 429 TTCTCCACAGAGTTCTCCAGAAAGGCTATCTTGGC 467
      |||||||

RESULT 9
ABK34853/c
ID ABK34853 standard; cDNA: 4298 BP.
XX
AC ABK34853;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human cDNA for novel secreted protein, SEQ ID 622.
XX
XX Human: ss; gene: secreted protein; immune deficiency; viral infection;
KM bacterial infection; fungal infection; autoimmune disorder; burn;
KM rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
KM diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
KM Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
KM coagulation disorder; haemophilia; inflammatory disorder; ulcer;
KM tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
KM lymphoid cell deficiency.
XX
XX Homo sapiens.
XX
OS WO200177290-A2.
XX
PN 18-OCT-2001.
XX
PD 29-MAR-2001; 2001WO-US10295.
XX
PR 06-APR-2000; 2000US-194941P.
XX
PA (GENY ) GENETICS INST INC.
XX
PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gutukota K, Graham JR;
XX
DR WPI: 2002-179323/23.
XX
XX Six hundred and twenty five polynucleotides derived from a variety of
XX human tissue sources which encode secreted proteins, useful for
XX treating immune deficiencies and disorders such as autoimmune disorders
XX
PT Claim 1: Page 336-337; 339pp; English.
XX
XX The invention relates to 625 polynucleotides which have been derived from
XX a variety of human tissue sources and sequences that hybridise to them.
XX proteins, their complements and sequences that hybridise to them.
XX Also included are a vector comprising the polynucleotide, a host cell
XX transformed with the vector, the proteins encoded by the
XX polynucleotides, antibodies that bind to the proteins and identification
XX of modulators of the proteins or the expression of the polynucleotide.
XX

```

CC The polynucleotides can be used as probes for the identification  
CC and isolation of full length cDNA and genomic DNA. The polynucleotides  
CC and proteins can also be used as nutritional supplements. The protein  
CC is useful in the treatment of various immune deficiencies and disorders  
CC such as viral infections, bacterial infections, fungal infections,  
CC autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis,  
CC autoimmune thyroiditis and diabetes) and allergic reactions and  
CC conditions (e.g. asthma). They are also useful for treating  
CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's  
CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),  
CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also  
CC useful for tissue regeneration, for wound healing and in the treatment  
CC of burns, incisions and ulcers. The proteins are also useful for  
CC regulating haematopoiesis, for treating myeloid or lymphoid cell  
CC deficiencies. The present sequence is one of the 625 cDNA sequences  
CC encoding a secreted protein.

SQ Sequence 4298 BP; 1258 A; 925 C; 941 G; 1174 T; 0 other;

Query Match	6.28;	Score	404.8;	DB	24;	Length	4298;
Best Local Similarity	99.58;	Pred. No.	4.7e-91;				
Matches	406;	Conservative	0;	Mismatches	2;	Indels	0;
				Gaps			0

QY	1710	TACATCATTTGGGATTTGGCTGACGTCGAAGATGACAGCATGTGAGATTAGGACAGCA	1769
Db	3548	TTTCAGTATATTTGGGATTTGGCTGACGTCGAAGATGACAGCATGTGAGATTAGGACAGCA	3488
QY	1770	AAGCAGATCCCAACTGCGATTTGCCCTGTGACGTGAGAACCTTATCCAGTAACTCGATTTA	1828
Db	3488	AAGCAGATCCCAACTGCGATTTGCCCTGTGACGTGAGAACCTTATCCAGTAACTCGATTTA	3422
QY	1830	TTTGACATCATCATTCGCAATTTTAAAGACTTCAGTGGCTACTTCCTGACTTGGCAGATCAAGTG	1888
Db	3428	TTTGACATCATCATTCGCAATTTTAAAGACTTCAGTGGCTACTTCCTGACTTGGCAGATCAAGTG	3363
QY	1890	CTAAGGCTTTTTAAAGCGTGATGACGCAAGCGCGCTACATATGATGACGTAAAGACACTGCA	1948
Db	3368	CTAAGGCTTTTTAAAGCGTGATGACGCAAGCGCGCTACATATGATGACGTAAAGACACTGCA	3309
QY	1950	GCAAGAAGATGCTCATTCAGGCTATCAGGAGATTTCGTTACTGACCCCGGATGCA	2003
Db	3308	GCAAGAAGATGCTCATTCAGGCTATCAGGAGATTTCGTTACTGACCCCGGATGCA	3249
QY	2010	TATTCACATATGTAGGCTCTGTGCACACCTGAGGGAGTATCAACAAGAAGACTTCCA	2066
Db	3248	TATTCACATATGTAGGCTCTGTGCACACCTGAGGGAGTATCAACAAGAAGACTTCCA	3183
QY	2070	GATAGGCTTCCAAACTGTGAGACAGAAATCAACATGAGGGAAGGAT	2117
Db	3188	GATAGGCTTCCAAACTGTGAGACAGAAATCAACATGAGGGAAGGAT	3141

RESULT 10  
ABL38290/c  
ID ABL38290 standard; cDNA; 402 BP.

AC ABL382907

DT 08-APR-2002 (first entry)

DE Human colon tumour antigen polynucleotide SEQ ID NO:1879

**KM** Human, colon cancer; colon tumour antigen; cytostatic; vaccine.  
**KW** colon tumour metastatic antigen; diagnosis; gene; ss.

OS Homo sapiens.

PN W0200196388-A2.

PD 20-DEC-2001

PF 08-JUN-2001; 2001WO-US18557

PR 09-JUN-2000; 2000US-210899P.  
PR 20-FEB-2001; 2001US-270216P.

PA (CORI-) CORIXA CORP

PI Jiang Y, Harlocker SL, Secrist H,

DR WPI; 2002-114514/15

PT Novel isolated colon tumor polynucleotide differentially expressed in  
PT colon tumor or colon metastatic tumor and polypeptides encoded by them  
PT useful for inhibiting development of cancer in patient -

PS Claim 1; SEQ ID 1879; 105pp; English.

CC ABI361612 to ABI38645 represent human colon tumour antigen cDNA clones (1)  
CC which were isolated from human colon tumour and colon metastatic tumour  
CC cDNA libraries. (1) have cytoskeletal activity and can be used in vaccine  
CC production. (1) can be used for stimulating and/or expanding T cells  
CC specific for a tumour protein on contact with the T cells. They are also  
CC useful for inhibiting the development of cancer in a patient. (1) can be  
CC used as probes or primers for nucleic acid hybridisation, for preparing  
CC mutant species primers, or primers for use in genetic constructions. (1)  
CC can be used in the diagnosis of a colon tumour.

50 Sequence 402 BP; 150 A; 67 C; 62 G; 121 T; 2 other;

Query Match	6.1%	Score	399.2	DB	24	Length	402
Best Local Similarity	99.8%	Pred. NO.	3.5e-90				
Matches	399	Conservative	1	Mismatches	0	Indels	0
						Gaps	0

QY	5360	TACTCAATGTGAGAGACTCTGGTCTGGATTTTTCATTAATTCACCTGATCAT	6019
Db	400	HACICCAANTGTGAGAGACTCTGGTCTGGATTTTTCATTAATTCACCTGATCAT	341
QY	6020	ATTGATCAGTAGATTAACGTAATAATAGCTTCAAAATTTTAAAGTGAATGTCAGTGTTTT	6079
Db	340	ATTGATCAGTAGATTAACGTAATAATAGCTTCAAAATTTTAAAGTGAATGTCAGTGTTTT	281
QY	6080	TCACGTATCAAAACAAATGTCAGGCTTTATTTAATAATCTCTCTGATTCATGGCATT	6133
Db	280	TCACGTATCAAAACAAATGTCAGGCTTTATTTAATAATCTCTCTGATTCATGGCATT	221
QY	6140	GTCCTACTGCTTATTTACATTTGTCACAAATTTATGCAATTTTACATGAAATATGCATTA	6199
Db	220	GTCCTACTGCTTATTTACATTTGTCACAAATTTATGCAATTTTACATGAAATATGCATTA	161
QY	6200	TTTGCCAGTTTTATTTATATAGCGCTATGAGCCCTCATGTGCATATAGAAACAGAAATCTA	6259
Db	160	TTTGCCAGTTTTATTTATATAGCGCTATGAGCCCTCATGTGCATATAGAAACAGAAATCTA	101
QY	6260	GCTTACCACCAAGTTGCCAAATGTTATCTAAGCAATTAAGTAATTTGTAGACATAGGACT	6319
Db	100	GCTTACCACCAAGTTGCCAAATGTTATCTAAGCAATTAAGTAATTTGTAGACATAGGACT	41
QY	6320	GCTAATCGACGTTGCGTCGTGATAGTCAATGACGATGT	6359
Db	40	GCTAATCGACGTTGCGTCGTGATAGTCAATGACGATGT 1	

RESULT 11  
AAA59384  
ID AAA59384 standard; DNA; 801 BP

AC AAA59384

DT 07-NOV-2000 (first entry)

DE DNA clone encoding a murine guanine nucleotide releasing factor 4

KW guanine nucleotide releasing factor 4; GRF-4; Ras activator;

KW cell differentiation; cell transformation; neuronal disorder; ss



```

XX Mus sp.
XX Key Location/Qualifiers
XX CDS 1..801
FT /*tag= a
FT "contains 5 internal stop codons and encodes
FT AAB07793; no termination codon given"
FT CDS 2..799
FT /*tag= b
FT "contains 8 internal stop codons and encodes
FT AAB07794; no termination codon given"
FT CDS 3..800
FT /*tag= c
FT "contains 5 internal stop codons and encodes
FT AAB07795; no termination codon given"
XX WO20043510-A2.
XX 27-JUL-2000.
XX 20-JAN-2000; 2000MO-CA00042.
XX 20-JAN-1999; 99CA-2259830.
XX (HSCR-) HSC RES & DEV LP.
XX Rolin D, Pham N;
XX WPI: 2000-499228/44.
XX P-PSDB; AAB07793, AAB07794, AAB07795.
XX Nucleic acids encoding guanine nucleotide releasing factor-4 useful for
XX the treatment of cancers and neuronal disorders.
XX Disclosure: Flg 19B; 89pp; English.
XX The present sequence encodes a murine guanine nucleotide releasing
XX factor (GRF)-4 (Ras activator) polypeptide. GRF4 activates Ras both
XX in vitro and in vivo. It directly binds cyclic adenosine monophosphate
XX (cAMP) directly via its cAMP-BD (cAMP/guanine monophosphate (cGM)
XX binding domain). GRF4 directly connects cAMP-generating (e.g. G protein
XX coupled receptors) or cGM-generating pathways to Ras. GRF4 activates
XX Ras in response to elevation of intracellular cAMP and/or cGM. GRF4
XX is a target for Nedd4 ubiquitination as it binds Nedd4. Activation of
XX the Ras signalling pathway controls numerous cellular functions, such as
XX cell metabolism, proliferation, differentiation and transformation.
XX Therefore modulation of Ras activity may provide a mechanism for
XX controlling diseases. GRF4 polynucleotides and polypeptides may be used
XX in the treatment of diseases associated with inappropriate GRF4
XX expression and activity such as cancers and neuronal disorders. The
XX GRF4 polypeptides may be used as antigens in the production of
XX antibodies against GRF4 and in assays to identify modulators (agonists
XX and antagonists) of GRF4 expression and activity. The anti-GRF4
XX antibodies and GRF4 antagonists may also be used to down regulate GRF4
XX expression and activity. Inhibition of Ras can reduce celluiose
XX proliferation and cancers.
XX Sequence 801 BP; 173 A; 266 C; 232 G; 130 T; 0 other:
XX
XX Query Match 5.9%; Score 389.2; DB 21; Length 801;
XX Best Local Similarity 74.6%; Pred. No. 1.7e-87;
XX Matches 586; Conservative 0; Mismatches 183; Indels 17; Gaps 7;
QY 4034 AACCTGGGCTCTTCCACAGATTACTGGGAGAGACTGAGAAAGTGCACAGCGCAAT 4093
Db 15 AACCTGAGCTCCACCGGCTGCGCCGCTCTAGAACTAGTGGATCCCCGGGCTGCAG 74
QY 4094 AAGCGGAGGGGTGAAGGATTTTCATTGAAGCCGAAGCAGTACGCTTGT 4153
Db 75 GAATTCAAGCGGTGGAGAGATCTCCGCTAGAGCAGACAGACATGGTGCCT 134
QY 4154 GACTACGAGAAACCAAGCCTGTCCCATGCTGCCACATAGCTGTGGCATCAAGTAC 4213

```

```

Db 135 GACTACGAGAGAGAGCAAACTGTCTCCATGCCGACATAGCTGTGACGCCGAGCAC 194
QY 4214 TACAAGGGGCTCATTTGCACGAAAGAGGCGAGTATTCAGAGAGCCCCGCCACCTCC 4273
Db 195 TACCAAGGAGCATCATCGACGAAAGAGGCGAGTACCGGGAGCGGCTCCACACTCC 254
QY 4274 CGGCTACATTGGAAATTTCCATTACTGACTTTCCAGAAAGGGCATCCCATCAGCAGAA 4333
Db 255 AGGCTACGTGGGCAATCCCATTTGCCATTTCCAGAAAGGGCTTTGCCACCGGCAAGAA 314
QY 4334 ACCGCGGAGCTACAAAGCTGGCCCTTCAGAGATCGGGATGGTGCAGATCTTCGACAC 4393
Db 315 GCCCGCGATTTACAAACGATGGCCCTCAGACGGTCCGATGGTGGCAGGGCCACTGAGC 374
QY 4394 AGCTGGGCTTATCTCCGTAACAGACAGCCACATGGGATATCCACACGACGAGCGTGGAA 4453
Db 375 CCCGGCAC-----GGGCAAGACCGCGCTGACCGCCAGCCAGCCGGCGGAG 425
QY 4454 CAACCTCAGTGCATTAACCGAACGAGTGCACCGCGCTCGCCCTTATCAGTCCCA 4513
Db 426 CAAGCCACAGTGGCACACAGCCACGACGACGACAGCCAGCGCTTCGACGCGCA 485
QY 4514 AGGCTTTTCCA--CCGAGAGGATGATGAACAAGTTTCTGCTGTTTGAGGCACAGAC 4571
Db 486 GGCTTCGAGAGCGGAGGAGAGCAAGATGAACAAGTGTGCTGTTTGAGGCGCAGGC 545
QY 4572 TTTTTCGAGAGCAGAGCCGACCTGAAAGAGCAGCAAGAGAGCTCTGATTCG 4631
Db 546 --TCTTGTATCCACAGTGAAGCC--CANAGAGAGCAGCAAGAGAGCTCTG 602
QY 4632 GAGCTTGGAACTCACATTTCTGAGAGCGGTGACAGCATTTGCTCTCTCTCTTAA 4691
Db 603 GAGCTTGGACAGCACA-TCTGAGATGCTGACAGCATTTGCTCTCTCTCTTAA 660
QY 4692 AGCAGCATGGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4751
Db 661 AGCAGCATGGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
QY 4752 GAAATTTGCCCTGCACATTTTCAGACTTTGTTGTAATGACAGTGCAGCATCTTCG 4811
Db 721 GAATTTGCCCTGCACATTTTCAGACTTTGTTGTAATGACAGTGCAGCATCTTCG 779
QY 4812 AGCTCC 4817
Db 780 AGCTGC 785

```

RESULT 12  
ABL37607/c  
ID ABL37607 standard; cDNA; 403 BP.  
XX ABL37607:  
AC 08-APR-2002 (first entry)  
DT 08-APR-2002 (first entry)  
XX Human colon tumour antigen polynucleotide SEQ ID NO:1196.  
DE Human colon tumour antigen polynucleotide SEQ ID NO:1196.  
KW Human; colon cancer; colon tumour antigen; cytostatic; vaccine;  
KW colon tumour metastatic antigen; diagnosis; gene; ss.  
XX Homo sapiens.  
XX WO200196388-A2.  
XX 20-DEC-2001.  
XX 08-JUN-2001; 2001WO-US18557.  
XX 09-JUN-2000; 2000US-210899P.  
XX 20-FEB-2001; 2001US-270216P.  
XX (CORI-) CORIXA CORP.









**BLANK PAGE**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 08:55:54 ; Search time 15987 Seconds  
(without alignments)  
11956.416 Million cell updates/sec

Title: US-09-911-826a-1  
Perfect score: 6568  
Sequence: 1 ctggcaccgtgagagattg.....aaaataataataatacat 6568

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hgt:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_pro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hgt\_hum:\*  
31: em\_hgt\_inv:\*  
32: em\_hgt\_other:\*  
33: em\_hgt\_mus:\*  
34: em\_hgt\_pin:\*  
35: em\_hgt\_rtd:\*  
36: em\_hgt\_mam:\*  
37: em\_hgt\_vrt:\*  
38: em\_sy:\*  
39: em\_hgtgo\_hum:\*  
40: em\_hgtgo\_mus:\*  
41: em\_hgtgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6568	100.0	6568	6	AX028128	AX028128 Sequence
2	6568	100.0	6568	6	AX408888	AX408888 Sequence
3	6568	100.0	6568	9	AB002311	AB002311 Human MRN
4	2011	30.6	160185	9	AC074344	AC074344 Homo sapi
5	1612	24.5	5742	9	AF394782	AF394782 Homo sapi
6	1584.4	24.1	6390	9	AF478567	AF478567 Homo sapi
7	1584.4	24.1	6401	9	AF117947	AF117947 Homo sapi
8	1581.4	24.1	1803	9	AF070570	AF070570 Homo sapi
9	1577.6	24.0	4806	9	AF478468	AF478468 Homo sapi
10	1502.6	22.9	4176	9	AF478469	AF478469 Homo sapi
11	1348.6	20.5	181796	2	AC124358	AC124358 Mus muscu
12	1341.4	20.4	3450	9	AK074318	AK074318 Homo sapi
13	1290	19.6	185108	2	AC118835	AC118835 Rattus no
14	863.4	13.1	194590	2	AL645962	AL645962 Mus muscu
15	845.6	12.9	148466	2	AC023167	AC023167 Mus muscu
16	728.6	11.1	195471	2	AC124186	AC124186 Mus muscu
17	716.2	10.9	173519	2	AC130840	AC130840 Mus muscu
18	697.2	10.6	3596	10	BC025553	BC025553 Mus muscu
19	601.4	9.2	2200	9	AK056286	AK056286 Homo sapi
20	560.8	8.5	51067	2	AC100578	AC100578 Mus muscu
21	399.2	6.1	402	6	AX341632	AX341632 Sequence
22	394.8	6.0	799	6	AX028130	AX028130 Sequence
23	390.4	5.9	4111	3	AF308449	AF308449 Caenorhab
24	390.4	5.9	4413	3	AF170796	AF170796 Caenorhab
25	390.4	5.9	4413	3	AF308447	AF308447 Caenorhab
26	390.4	5.9	4413	3	AF308448	AF308448 Caenorhab
27	388.2	5.9	403	6	AX340949	AX340949 Sequence
28	336.4	5.1	185108	2	AC118835	AC118835 Rattus no
29	333.2	5.1	181796	2	AC124358	AC124358 Mus muscu
30	326.4	5.0	3670	9	HSMB00982	AL117397 Homo sapi
31	289	4.4	76375	2	AC017591	AC017591 Drosophill
32	289	4.4	209071	3	AC005285	AC005285 Drosophill
33	289	4.4	214622	3	AC093100	AC093100 Drosophill
34	289	4.4	261872	3	AE003613	AE003613 Drosophill
35	287.4	4.4	5025	3	AF434686	AF434686 Drosophill
36	230.8	3.5	192912	10	AL607091	AL607091 Mouse DNA
37	227.8	3.5	263	6	AX153664	AX153664 Sequence
38	219.6	3.3	85181	9	AC004622	AC004622 Homo sapi
39	219.4	3.3	108424	10	AL669827	AL669827 Mouse DNA
40	208.2	3.2	38799	3	CEK04D7	Z69664 Caenorhabd
41	198	3.0	2675	10	BC019702	BC019702 Mus muscu
42	174.4	2.7	110523	2	AC103244	AC103244 Rattus no
43	152.2	2.3	158	6	AX072819	AX072819 Sequence
44	150	2.3	133255	2	AC121410	AC121410 Rattus no
45	139.2	2.1	51067	2	AC100578	AC100578 Mus muscu

ALIGNMENTS

RESULT 1	LOCUS	SEQUENCE	AX028128	6568 bp	DNA	1 linear	PAT 16-SRP-2000
AX028128	DEFINITION	Sequence 1 from Patent WO0043510.					
AX028128	ACCESSION	AX028128					
AX028128	VERSION	AX028128.1	GI:10189937				
	KEYWORDS						
	SOURCE	human.					
	ORGANISM	Homo sapiens					
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
	REFERENCE	1 (bases 1 to 6568)					
	AUTHORS	Rotin,D. and Pham,N.					
	TITLE	Ras activator nucleic acid molecules, polypeptides and methods of use					

JOURNAL Patent: WO 0043510-A 1 27-JUL-2000;  
ROTIN DANIELA (CA) ; HSC RES DEV LP (CA) ; PHAM NAM (CA)  
FEATURES Location/Qualifiers  
source 1..6568  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
63..4562  
CDS  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC09161.1"  
/db\_xref="GI:1018938"  
/translation="MKPLAPAHGVMDGOEKHSLPADFTKLHLDLSLPHOYTHVSS  
HSCSTISDSSSSLSDIQATHESEAGMDLSGLPTAVSDSDDEEDIEFASDPLM  
SRDVIDCLEKDDIDRTDDTEQLFEHQLPFPAMTMSVRELCAVAFVAVERAG  
TYLVNDEELDSWVTLNGSVETPDGAEILCWNSGVSPTMDKEKMGKVRKV  
DQCFVCIADQDCIRILNQEKRMQVVEEGEILVYKRELDRTGRGHVYIKGTS  
ERLTMHVEBHSVDPDTEFLTYRTFLSMEYKGLLEFNPSLRDVTYRVL  
LWVNHNPDEEDPAMTRFLEEFENLBERKMGHLRLINICAAKARMLTTPRS  
REAPLPFILGSEKGFIVDSVDSKATKRGQIILEVNGNFENITOLSKAM  
EILRNTHLSTVTKNLVFKELTLRLEKRNKGFHLKIGDIKASKYSTIPDLAV  
VEQVIGLEKVNKSKANTVGRNKLKILDKTRIILPKRPYMDIGISODSTVGL  
ROTKHPTALPVSGTSSNPDLQSHHRLDPSAPDLQVLRFKADQOSRYIMI  
SKDTAKEVVIOAREFAVATPDQYSLCEVSTPEGVTKORLPOLSKLADRIOLS  
GRYILKNNMETELGDEDOELRESQISLQSTVEVATOLSMNFELRNEPTE  
YIDILFKLRKSTCANLKRREYINQETEWASEILREINOLKRMKILHFIKILALHC  
RECKNRSMAHIIISGLNLAFLVARIKTTMELPKRYKELFODLDPSPRNMAKYVN  
LNSQNLQPTPLLPVYKDLTLRHNQSDKVDGLVNEFKRLAKETIRHVGMAKVN  
MDPLMFRTRKRRKRSLSQSTNAVLDVAQGTGKKRVRSSFLLNAKLYEDAO  
MARKVQYLSNLELMEDESLQTLSDCEPATITLKNPDKPVSESPVAPRAGS  
QOKAOSLPQOQPPPAHKINOGLQVAVSLYSPKRPVKDLPFGINSPOALKLIL  
SLSEGLERHKROAEDTISNASOLSPTSPSSPRKGYTLAPGVDNFDSGHS  
ELSRSTVSNSSPDSVPVSLHDERQRHVSIVENLNGMRMTELEPDYISGS  
YAPMSERGLIATATVYSSTFELSQDQDRASLDAAQSGRSMTSKSGSHDNIOT  
IQQORSKETLPFGHTHPDYSGDPAWMASSMDQIMFSDHSTRKYRMONSRESLQQA  
QOSAWASSTGYWGEDSEGDGTIKRGGADVSIIESSSLVITTELRKVPMPHI  
AVASSTTGLIARKEGRREPPTPEYDIPITDPEGSHAPRPDPVNALQSR  
MVARSPTAGPSSVQOPHGHPTSSRPVKNPQMKPMSDPRLAPYOSQGFSTEDEDE  
OVSAY"

BASE COUNT 1974 a 1400 c 1463 g 1731 t  
ORIGIN  
Query Match 100.0%; Score 6568; DB 6; Length 6568;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTGCCATCGTAGAGATTGGTACATGCTGTAAATTCAGTACGATATGTTCTTCA 60  
DB 1 CTTGCCATCGTAGAGATTGGTACATGATGTAAATTCAGTACGATATGTTCTTCA 60  
QY 61 TTATGAACCACTAGCAATCCAGCTAACCAATGAGATTGGCCAGCAGGAAACACT 120  
DB 61 TTATGAACCACTAGCAATCCAGCTAACCAATGAGATTGGCCAGCAGGAAACACT 120  
QY 121 CACTTCCTGAGATTTCACAAAACTGATCTTACTGACAGTCTCCACACAGGTGACCC 180  
DB 121 CACTTCCTGAGATTTCACAAAACTGATCTTACTGACAGTCTCCACACAGGTGACCC 180  
QY 121 CACTTCCTGAGATTTCACAAAACTGATCTTACTGACAGTCTCCACACAGGTGACCC 180  
DB 121 CACTTCCTGAGATTTCACAAAACTGATCTTACTGACAGTCTCCACACAGGTGACCC 180  
QY 131 AGCTTTCTTCTAGCCATTACAGATGTACTATCAGTATGTTGGGAGCAGCAGTCTTT 240  
DB 131 AGCTTTCTTCTAGCCATTACAGATGTACTATCAGTATGTTGGGAGCAGCAGTCTTT 240  
QY 181 AGCTTTCTTCTAGCCATTACAGATGTACTATCAGTATGTTGGGAGCAGCAGTCTTT 240  
DB 181 AGCTTTCTTCTAGCCATTACAGATGTACTATCAGTATGTTGGGAGCAGCAGTCTTT 240  
QY 241 CTGATATCTACACAGCCACAGAAACGAGGCTGTGATATGAGACTTGATGGGTTGCCAG 300  
DB 241 CTGATATCTACACAGCCACAGAAACGAGGCTGTGATATGAGACTTGATGGGTTGCCAG 300  
QY 301 AAACAGAGTGGATTCCGAAGACGACGAGATGAAGAAGACATTGAGAGATCAGATC 360  
DB 301 AAACAGAGTGGATTCCGAAGACGACGAGATGAAGAAGACATTGAGAGATCAGATC 360  
QY 361 CTCTGATGAGCAGGACATGTGTAGAGACTGCTAGAGAAGACCAATGTGACCGGACAG 420  
DB 361 CTCTGATGAGCAGGACATGTGTAGAGACTGCTAGAGAAGACCAATGTGACCGGACAG 420

QY 421 ATGATGACATGAGAACACTCTTGAAATTTATGCACCACTGGCTCTTGGCAATATGA 480  
DB 421 ATGATGACATGAGAACACTCTTGGAAATTTATGCACCACTGGCTCTTGGCAATATGA 480  
QY 481 CAATGTCAGTGAAGCGAGAACTCTGTGCTGTGATGATGCTGTCAGATGTTGAAAGCAG 540  
DB 481 CAATGTCAGTGAAGCGAGAACTCTGTGCTGTGATGATGCTGTCAGATGTTGAAAGCAG 540  
QY 541 GGACCATFAGTGTAAATGATGTGTGAAGACCTGACCTCGGTGACATGATGATGAT 600  
DB 541 GGACCATFAGTGTAAATGATGTGTGAAGACCTGACCTCGGTGACATGATGATGAT 600  
QY 601 CTGTGAGAGTGAATTTTCCAGATGGAAGAAACGAAATACGTGATGGAATTAATTTTG 660  
DB 601 CTGTGAGAGTGAATTTTCCAGATGGAAGAAACGAAATACGTGATGGAATTAATTTTG 660  
QY 661 GTGTCTCTCTACCATGAGCAAAAGAAATACATGAAGAGTATGAGAACAGTGTGATG 720  
DB 661 GTGTCTCTCTACCATGAGCAAAAGAAATACATGAAGAGTATGAGAACAGTGTGATG 720  
QY 721 ACTGCCAGTTTGTCTGCATAGCCACAGATTTAGTCCGCTATTCATCAATCAAGTAGAA 780  
DB 721 ACTGCCAGTTTGTCTGCATAGCCACAGATTTAGTCCGCTATTCATCAATCAAGTAGAA 780  
QY 781 AGAACATGCAAAAGTTGAAAGAGAGAGAGATTTGTATGTGTGAAGAAACACGAGAAC 840  
DB 781 AGAACATGCAAAAGTTGAAAGAGAGAGAGATTTGTATGTGTGAAGAAACACGAGAAC 840  
QY 841 TTGATGAACTGACACAAAGAAAGGACACATTTGTCATCAAGGTTACTGAGAAAGTTTA 900  
DB 841 TTGATGAACTGACACAAAGAAAGGACACATTTGTCATCAAGGTTACTGAGAAAGTTTA 900  
QY 901 CAATGCAATTTGGTGGAGAGACATTCAGTATGATCCACATTCATAGAAAGACTTCTGT 960  
DB 901 CAATGCAATTTGGTGGAGAGACATTCAGTATGATCCACATTCATAGAAAGACTTCTGT 960  
QY 961 TGACCTATAGGACTTTTCTTCTTCTAGCCCAATGGAATGGGCAAAAGTTATGGAGTGT 1020  
DB 961 TGACCTATAGGACTTTTCTTCTTCTAGCCCAATGGAATGGGCAAAAGTTATGGAGTGT 1020  
QY 1021 TTATATGCCCGGACCTCAGGATAGGTTTACAGGGTAGATATTTGGGATTAATATATC 1080  
DB 1021 TTATATGCCCGGACCTCAGGATAGGTTTACAGGGTAGATATTTGGGATTAATATATC 1080  
QY 1081 ACTTCATGACTTTGAAAGGAGATCTGCAATGACTGATTTTGAAGAATTTGAAGAAACA 1140  
DB 1081 ACTTCATGACTTTGAAAGGAGATCTGCAATGACTGATTTTGAAGAATTTGAAGAAACA 1140  
QY 1141 ATCTGAAAGAGAGAAATATGGTGGACACTTAAGGCTGTGATATTCGGTGTGCTTA 1200  
DB 1141 ATCTGAAAGAGAGAAATATGGTGGACACTTAAGGCTGTGATATTCGGTGTGCTTA 1200  
QY 1201 AAGCAAAAGAGATGTATGATGATCAAAACCAATCCGAGAGAGCTCTTTCCTTTTA 1260  
DB 1201 AAGCAAAAGAGATGTATGATGATCAAAACCAATCCGAGAGAGCTCTTTCCTTTTA 1260  
QY 1261 TCTTACTTGGAGGCTCTGAGAAAGGATTTGAAATCTTTTGTGACAGTGTAGATTCA 1320  
DB 1261 TCTTACTTGGAGGCTCTGAGAAAGGATTTGAAATCTTTTGTGACAGTGTAGATTCA 1320  
QY 1321 GCAAGACACTGAAAGCGCTTGAAGAGGGGGATCGATATTAAGAAATTAAGAGGCGCAA 1380  
DB 1321 GCAAGACACTGAAAGCGCTTGAAGAGGGGGATCGATATTAAGAAATTAAGAGGCGCAA 1380  
QY 1381 ACTTGAAGACATTCACAGTGCAGAAACCTATGGAATTTTGAAGAAATTAACACATTTAT 1440  
DB 1381 ACTTGAAGACATTCACAGTGCAGAAACCTATGGAATTTTGAAGAAATTAACACATTTAT 1440  
QY 1441 CTATCAGTGTGAAACCAATTTATTTGTATTTAAAGAACTTCAACAAGATTTTCAGAA 1500  
DB 1441 CTATCAGTGTGAAACCAATTTATTTGTATTTAAAGAACTTCAACAAGATTTTCAGAA 1500  
QY 1501 AGAAAGAAATGTGTCCTCCACCTTCCCTTAATTTGTTGATGATTTAAAGGCCAGTCCCT 1560

1501 AAAAAAGAAATGTGCCCCACCTCTCTAAATGGTGACATTTAAAAAGGCGAGTCGT 1560  
QY 1561 ACTCCATTCCAAATCTGTGTAGATGTGAACAGGTGTAGGACTTGA AAAAGTGAACA 1620  
DB 1561 ACTCATTCACAAATCTGTGTAGATGTGAACAGGTGTAGGACTTGA AAAAGTGAACA 1620  
QY 1621 AAAAAAGTAAAGCAACACTGTGGGAGGAAGAACAGGTGA AAAAGTGAACACTGACACA 1680  
DB 1621 AAAAAAGTAAAGCAACACTGTGGGAGGAAGAACAGGTGA AAAAGTGAACACTGACACA 1680  
QY 1681 CTGGATCAGTATCTTGGCCACAGAAAACCATACATGATATTGGGATTTGGCTCAGTCAAG 1740  
DB 1681 CTGGATCAGTATCTTGGCCACAGAAAACCATACATGATATTGGGATTTGGCTCAGTCAAG 1740  
QY 1741 ATGACAGCATTAAGATTAAGGACAGAACACATCCCACTGATGCTGTCAAGT 1800  
DB 1741 ATGACAGCATTAAGATTAAGGACAGAACACATCCCACTGATGCTGTCAAGT 1800  
QY 1801 GAACCTATATCATCAGTATCTGATTTATGTCAGTCACATCATGCTATTTAGACTTCA 1860  
DB 1801 GAACCTATATCATCAGTATCTGATTTATGTCAGTCACATCATGCTATTTAGACTTCA 1860  
QY 1861 GTGCTACTCTGATCTTGCAGATCAAGTCTAAGGTTTTTAAGGCTGATCAGCAAAAGCC 1920  
DB 1861 GTGCTACTCTGATCTTGCAGATCAAGTCTAAGGTTTTTAAGGCTGATCAGCAAAAGCC 1920  
QY 1921 GCTACATCATGATCAGTAAAGACACTACAGCAAGAAAGTGTGCTACAGGCTATAGGG 1980  
DB 1921 GCTACATCATGATCAGTAAAGACACTACAGCAAGAAAGTGTGCTACAGGCTATAGGG 1980  
QY 1981 AGTTGCTCTGATCTGACCCGATCAATATTCACTATGTGAGGCTGTGTCACACTG 2040  
DB 1981 AGTTGCTCTGATCTGACCCGATCAATATTCACTATGTGAGGCTGTGTCACACTG 2040  
QY 2041 AGGAGTAAATCAAAACAAAGAGATTCAGATCAGCTTTCCAAACTTGCAGACAGAAATAC 2100  
DB 2041 AGGAGTAAATCAAAACAAAGAGATTCAGATCAGCTTTCCAAACTTGCAGACAGAAATAC 2100  
QY 2101 AACTGAGTGAAGTATTTATTTAAACATGAGTAAACACATGTAACAACTCTTTGTCAGATG 2160  
DB 2101 AACTGAGTGAAGTATTTATTTAAACATGAGTAAACACATGTAACAACTCTTTGTCAGATG 2160  
QY 2161 AAGATGCTCAGAGATTTGAGAGAGATCAAAATTTCCCTCTTACGCTCAGCAGCTGTG 2220  
DB 2161 AAGATGCTCAGAGATTTGAGAGAGATCAAAATTTCCCTCTTACGCTCAGCAGCTGTG 2220  
QY 2221 AAGTTGCAACACAGCTCTATGCGAAATTTGAACTTTGCAACATTTGAACCTACTG 2280  
DB 2221 AAGTTGCAACACAGCTCTATGCGAAATTTGAACTTTGCAACATTTGAACCTACTG 2280  
QY 2281 AATATATAGATGTTTATTTAAACATGAGTAAACACAGCTGTGCAACCTGAAGAGAT 2340  
DB 2281 AATATATAGATGTTTATTTAAACATGAGTAAACACAGCTGTGCAACCTGAAGAGAT 2340  
QY 2341 TTGAAGAAGTCAATTAACAGAAACATTTTGGGTAGCATCTGAAATTTCCAGGAAACAA 2400  
DB 2341 TTGAAGAAGTCAATTAACAGAAACATTTTGGGTAGCATCTGAAATTTCCAGGAAACAA 2400  
QY 2401 ACCAGCTGAAGAGATGAAGATCATTAAGCATTTCAATCAAGATAGCAGTCACTGAGG 2460  
DB 2401 ACCAGCTGAAGAGATGAAGATCATTAAGCATTTCAATCAAGATAGCAGTCACTGAGG 2460  
QY 2461 AATGCAAGAATTTTAACTCAATGTTTGAATCATCATGCTGACCTAAACCTGGACCCAGTGG 2520  
DB 2461 AATGCAAGAATTTTAACTCAATGTTTGAATCATCATGCTGACCTAAACCTGGACCCAGTGG 2520  
QY 2521 CAAGACTGGGAAGCAGCTGGAGAAACTTCCCAATTAATACGAAAAACTATTTCAAGATC 2580  
DB 2521 CAAGACTGGGAAGCAGCTGGAGAAACTTCCCAATTAATACGAAAAACTATTTCAAGATC 2580  
QY 2581 TCCAGACCTGTTTGAATCTTCCAGAAACATGCAAAATATGCTAATGCT 2640  
DB 2581 TCCAGACCTGTTTGAATCTTCCAGAAACATGCAAAATATGCTAATGCT 2640

2581 TCCAGACCTGTTTGAATCTTCCAGAAACATGCAAAATATGCTAATGCT 2640  
QY 2641 AAAATATACACCTCCCATATTCCTCTATTTCCCAATTTATCAAAAAGATCTCAGCTTCC 2700  
DB 2641 AAAATATACACCTCCCATATTCCTCTATTTCCCAATTTATCAAAAAGATCTCAGCTTCC 2700  
QY 2701 TTCAAGAGAAATGACTCAAAAGTGAAGGAGGCTGTCAATTTTGAAGCTAAGGATCA 2760  
DB 2701 TTCAAGAGAAATGACTCAAAAGTGAAGGAGGCTGTCAATTTTGAAGCTAAGGATCA 2760  
QY 2761 TTGCAAAAGAAATTTGTCAGTGTGGCCGAATGCTTCACTGAATGCAATGCCCTCA 2820  
DB 2761 TTGCAAAAGAAATTTGTCAGTGTGGCCGAATGCTTCACTGAATGCAATGCCCTCA 2820  
QY 2821 TGTTCAGACCTGGAGAAAGAAATGAGGAGATTTGGGGTCTCTACCCAGGAGTACAA 2880  
DB 2821 TGTTCAGACCTGGAGAAAGAAATGAGGAGATTTGGGGTCTCTACCCAGGAGTACAA 2880  
QY 2881 ATGCAACAGTGTAGATGTTGCTCAGACAGGTGTCATAAAAAGCGGATACGTAGTT 2940  
DB 2881 ATGCAACAGTGTAGATGTTGCTCAGACAGGTGTCATAAAAAGCGGATACGTAGTT 2940  
QY 2941 CCTTTCTCAATGCAAAAAAGCTTTATGAAGATGCCAAATGCTCGAAAAAGTGAACAGT 3000  
DB 2941 CCTTTCTCAATGCAAAAAAGCTTTATGAAGATGCCAAATGCTCGAAAAAGTGAACAGT 3000  
QY 3001 ACCTTTCCAAATTTGGAGCTAGAAATGAGAGAGAGTCTTCAGACATTTATCTCTCAGT 3060  
DB 3001 ACCTTTCCAAATTTGGAGCTAGAAATGAGAGAGAGTCTTCAGACATTTATCTCTCAGT 3060  
QY 3061 GTGAGCCAGCAACCAACATTTGCTAAGAAATCTGCTGCAAAAAAGCGCTGAATTCG 3120  
DB 3061 GTGAGCCAGCAACCAACATTTGCTAAGAAATCTGCTGCAAAAAAGCGCTGAATTCG 3120  
QY 3121 AGACCTCTCAGTACGCTCAAGGAGGCTCACACAGAAAGCTCAAGTCCCTGCCACAGC 3180  
DB 3121 AGACCTCTCAGTACGCTCAAGGAGGCTCACACAGAAAGCTCAAGTCCCTGCCACAGC 3180  
QY 3181 CCCAGCAGCAGCCACACACATTAATAATCAACAGGAGCTACAGGTTCCGCCGTGT 3240  
DB 3181 CCCAGCAGCAGCCACACACATTAATAATCAACAGGAGCTACAGGTTCCGCCGTGT 3240  
QY 3241 CCCTTTATCTTCACGAGAAAGAGTCCCGTAAGATCTCCCACTTTTGGCATAACT 3300  
DB 3241 CCCTTTATCTTCACGAGAAAGAGTCCCGTAAGATCTCCCACTTTTGGCATAACT 3300  
QY 3301 CTCACAGCTTTAAAAAAATTTCTTTGTCGTAAGAGAGTGTGGAACGTACACA 3360  
DB 3301 CTCACAGCTTTAAAAAAATTTCTTTGTCGTAAGAGAGTGTGGAACGTACACA 3360  
QY 3361 AGAAACAGGCTGAAGATACAAATTCAAATGCAATCTTCGAGGCTTCTCCCTCACTT 3420  
DB 3361 AGAAACAGGCTGAAGATACAAATTCAAATGCAATCTTCGAGGCTTCTCCCTCACTT 3420  
QY 3421 CTCACAGAGTTCCTCAAGGAAGGCTATCTTTGGCTCCCACTGTACTGTGATTAAT 3480  
DB 3421 CTCACAGAGTTCCTCAAGGAAGGCTATCTTTGGCTCCCACTGTACTGTGATTAAT 3480  
QY 3481 TTTCAATCTGTGTCAAGTGAATTTCTTCAAGATCAGTATTTGTAAGCAATTCGCTT 3540  
DB 3481 TTTCAATCTGTGTCAAGTGAATTTCTTCAAGATCAGTATTTGTAAGCAATTCGCTT 3540  
QY 3541 TTGACATCAGTGGCAGTCACTGACAGATGAGAGGCGCAGAGGCAATCTGTGAGAGTGC 3600  
DB 3541 TTGACATCAGTGGCAGTCACTGACAGATGAGAGGCGCAGAGGCAATCTGTGAGAGTGC 3600  
QY 3601 TTGAAACAAACCTAAGGATGAGCAGATGAGAGGCGGACATGATTTGAACCTGATCAGT 3660  
DB 3601 TTGAAACAAACCTAAGGATGAGCAGATGAGAGGCGGACATGATTTGAACCTGATCAGT 3660  
QY 3661 ATAGCTTGGGGTCTATGCAACAATGTCGAGGCGCAGGCTTATATGCTACAGCTACAG 3720  
DB 3661 ATAGCTTGGGGTCTATGCAACAATGTCGAGGCGCAGGCTTATATGCTACAGCTACAG 3720

Qy	3721	TAATTTCTTCTTCCAAGCACAGAGAACTTTCCAGAGATCAGGGGATCGCGGCTACTTG	3780
Db	3721	TAATTTCTTCTTCCAAGCACAGAGAACTTTCCAGAGATCAGGGGATCGCGGCTACTTG	3780
Qy	3781	ATGCGTCGACAGTGGCGGTGGACCTGGAGCTGCATGCTCAAGTGGGCTCCATGTAATA	3840
Db	3781	ATGCGTCGACAGTGGCGGTGGAGCTGGACCTCATGCTCAAGTGGGCTCCATGTAATA	3840
Qy	3841	TACAGCAGATCCAGCACAGAGAGCTGGGAGACTTTCATTCGGGCACTCACTTGG	3900
Db	3841	TACAGCAGATCCAGCACAGAGAGCTGGGAGACTTTCATTCGGGCACTCACTTGG	3900
Qy	3901	ATTATTCAGGGGATCCTGAGGTTTATGGGATTCAGAGCCATATGAGCAAAATATGT	3960
Db	3901	ATTATTCAGGGGATCCTGAGGTTTATGGGATTCAGAGCCATATGAGCAAAATATGT	3960
Qy	3961	TTTCTGATATGACACAAAGTATTAACAGGCCAAATTAACAAAGTAGAGAGACCTTGAACAAG	4020
Db	3961	TTTCTGATATGACACAAAGTATTAACAGGCCAAATTAACAAAGTAGAGAGACCTTGAACAAG	4020
Qy	4021	CCGAGTCCCGAGCAAGCTGGCGTCTTCCACAGGTACTGGGAGAGAACTCAGAAAGTG	4080
Db	4021	CCGAGTCCCGAGCAAGCTGGCGTCTTCCACAGGTACTGGGAGAGAACTCAGAAAGTG	4080
Qy	4081	ACACAGCACAATAAAGCGAGGGGTGGAAGAGATGTTTCCATGTAAGCCGAAGCAGTA	4140
Db	4081	ACACAGCACAATAAAGCGAGGGGTGGAAGAGATGTTTCCATGTAAGCCGAAGCAGTA	4140
Qy	4141	GCCTAAGCTGTGTGACTACAGGAAGAAACCAAGCTTCCCATGCTGCCCATATAGCTG	4200
Db	4141	GCCTAAGCTGTGTGACTACAGGAAGAAACCAAGCTTCCCATGCTGCCCATATAGCTG	4200
Qy	4201	TGGCATCAAGTACTACAAAGGGGCTCATTTGCAGAAAGAGGGGACAGTATCGAGAGCCCC	4260
Db	4201	TGGCATCAAGTACTACAAAGGGGCTCATTTGCAGAAAGAGGGGACAGTATCGAGAGCCCC	4260
Qy	4261	CGCCACACCCCTCCCGGCTACATTGGANTTCCATTACTGACTTTCCAGAAAGGCACTGCC	4320
Db	4261	CGCCACACCCCTCCCGGCTACATTGGANTTCCATTACTGACTTTCCAGAAAGGCACTGCC	4320
Qy	4321	ATCCAGCAGAGAAACGGCGGAGCTACAAAGCTGGCCGCTTCAGAAATCGCGGATGGTCGAC	4380
Db	4321	ATCCAGCAGAGAAACGGCGGAGCTACAAAGCTGGCCGCTTCAGAAATCGCGGATGGTCGAC	4380
Qy	4381	GATTCCTCCGACACAGCTGGGCTTCATCGTACAGCAGCCACATGGGCATCCACACGCA	4440
Db	4381	GATTCCTCCGACACAGCTGGGCTTCATCGTACAGCAGCCACATGGGCATCCACACGCA	4440
Qy	4441	GCAGGCGCTGGAAACCACTCAGTGGCATTAACCGAAGCAGTGCAGCCGCGCTGCACC	4500
Db	4441	GCAGGCGCTGGAAACCACTCAGTGGCATTAACCGAAGCAGTGCAGCCGCGCTGCACC	4500
Qy	4501	CTTATCAGTCCCAAGGGTTTTCCACGAGAGAGTGAAGATGAACAAGTTTCTGTGTTTT	4560
Db	4501	CTTATCAGTCCCAAGGGTTTTCCACGAGAGAGTGAAGATGAACAAGTTTCTGTGTTTT	4560
Qy	4561	GAGGCACAGACTTTTCTGGAAGACAGAGCCAGCCACTGAAAGAGAGACACAAAGAGAGCT	4620
Db	4561	GAGGCACAGACTTTTCTGGAAGACAGAGCCAGCCACTGAAAGAGAGACACAAAGAGAGCT	4620
Qy	4621	CCTGAGCATTTGGAGCCTTGGAACTCATTTCTGAGAGCGGTGACCAAGTTTGCTCTTCC	4680
Db	4621	CCTGAGCATTTGGAGCCTTGGAACTCATTTCTGAGAGCGGTGACCAAGTTTGCTCTTCC	4680
Qy	4681	CTTGCTTTAAAGCAGCATGGGGCTTCTCTCCCTTCTTCTTCCCTTTTGATGTGA	4740
Db	4681	CTTGCTTTAAAGCAGCATGGGGCTTCTCTCCCTTCTTCTTCCCTTTTGATGTGA	4740
Qy	4741	AATTACTGTGAAGAAATTTGCCCTGGCAGCTTTCAACACTTTTGTTGCTTGAATGCAAGTGC	4800
Db	4741	AATTACTGTGAAGAAATTTGCCCTGGCAGCTTTCAACACTTTTGTTGCTTGAATGCAAGTGC	4800

Qy	4801	AGCAATCTTGGAGCTCCCACTGTGTGCTGCTGCCACATACACACAGTATATCTCCAAATTC	4860
Db	4801	AGCAATCTTGGAGCTCCCACTGTGTGCTGCTGCCACATACACACAGTATATCTCCAAATTC	4860
Qy	4861	CAAGATCATCACAAACAAGATGATCTCCTGGCTGCACCTTCCAAATGCCGGAGAATTT	4920
Db	4861	CAAGATCATCACAAACAAGATGATCTCCTGGCTGCACCTTCCAAATGCCGGAGAATTT	4920
Qy	4921	TTTTTAATCTCTTTTAAATTTCAATCCAGTCTACACTTGTATCTCAATTGGGATATG	4980
Db	4921	TTTTTAATCTCTCTTTAAATTTCAATCCAGTCTACACTTGTATCTCAATTGGGATATG	4980
Qy	4981	AGAAAGCTAGACCATTAAGTACTCTTGGGGCTTTTAACTCCACCAAGAGAACAAAGAAAA	5040
Db	4981	AGAAAGCTAGACCATTAAGTACTCTTGGGGCTTTTAACTCCACCAAGAGAACAAAGAAAA	5040
Qy	5041	CAATGAATCTCTTTGAGTACAGTGTCTGTCTCACTGTTTAAATCAATGTCTCTTTAAAAA	5100
Db	5041	CAATGAATCTCTTTGAGTACAGTGTCTGTCTCACTGTTTAAATCAATGTCTCTTTAAAAA	5100
Qy	5101	AAAAAATGATTTAAAGATTTTGTCTAGAGAGTAATATATATCCATTTAAATGATTACAG	5160
Db	5101	AAAAAATGATTTAAAGATTTTGTCTAGAGAGTAATATATATCCATTTAAATGATTACAG	5160
Qy	5161	TATATATTTTAAACCTTAAGTAGGGTTGCCAGCCTGTGTTTCTGAAAAACCAATATGCCGG	5220
Db	5161	TATATATTTTAAACCTTAAGTAGGGTTGCCAGCCTGTGTTTCTGAAAAACCAATATGCCGG	5220
Qy	5221	ACAGGGTGTGGCCACACCAAGAAGAGGGGAAGACCTGTGTGAGACCTGGCTTCCCATG	5280
Db	5221	ACAGGGTGTGGCCACACCAAGAAGAGGGGAAGACCTGTGTGTGACCTGTGGCTTCCCATG	5280
Qy	5281	TCCCTTGGCTCACCCGCGAAGTGCCTTTCCTGGAAGTAAATAATGTTAGCCAAATTTAA	5340
Db	5281	TCCCTTGGCTCACCCGCGAAGTGCCTTTCCTGGAAGTAAATAATGTTAGCCAAATTTAA	5340
Qy	5341	TACCAAGACACCTCATCTGCTCCTTCCCACTGAGTATGGGTTCTTCTGTAAACTGTTCG	5400
Db	5341	TACCAAGACACCTCATCTGCTCCTTCCCACTGAGTATGGGTTCTTCTGTAAACTGTTCG	5400
Qy	5401	CACATGGCCAGGGGAGGAGACTAGACCTCTTGCTGCTGTACGCTTATGAGAGCAGG	5460
Db	5401	CACATGGCCAGGGGAGGAGACTAGACCTCTTGCTGCTGTACGCTTATGAGAGGAGG	5460
Qy	5461	ACGGTGTCAATGGCGAGTGTCTCTGCTCATTTGATGATGATGGAACCCCAATTTTTAA	5520
Db	5461	ACGGTGTCAATGGCGAGTGTCTCTCTCATTTGATGATGATGGAACCCCAATTTTTAA	5520
Qy	5521	GTATATTTCTCTTGATTTTGTTAATTAATTAGAGGTGTAGTTTGTTTTGTTTTGGT	5580
Db	5521	GTATATTTCTCTTGATTTTGTTAATTAATTAGAGGTGTAGTTTGTTTTGTTTTGGT	5580
Qy	5581	TTTTTTTTAGAGAAACATTTATAACTGATAGCAATTCAGTGAAGACGCTTGGGATGTT	5640
Db	5581	TTTTTTTTAGAGAAACATTTATAACTGATAGCAATTCAGTGAAGACGCTTGGGATGTT	5640
Qy	5641	GGACCTAATGCGACAGCTGTTTAACTGCTCTTTCACAGACGCTCCCTTAAATGAAATGGC	5700
Db	5641	GGACCTAATGCGACAGCTGTTTAACTGCTCTTTCACAGACGCTCCCTTAAATGAAATGGC	5700
Qy	5701	ATTAGGAATTAACAAAGCCTTTAAAGCTGATTAAGAATCAAAAACCTGTGTAGCATGCC	5760
Db	5701	ATTAGGAATTAACAAAGCCTTTAAAGCTGATTAAGAATCAAAAACCTGTGTAGCATGCC	5760
Qy	5761	AGCCTTTCGAAGCAGGTTAGTACACCAAAGCTAACTCCAAAGTGGCTTATATGAGACCTG	5820
Db	5761	AGCCTTTCGAAGCAGGTTAGTACACCAAAGCTAACTCCAAAGTGGCTTATATGAGACCTG	5820
Qy	5821	CATTTATGAGAGGCTTAAGTGTAGCAACCAATCTGCTACAGCTCTCTATTAACCTTAAAT	5880
Db	5821	CATTTATGAGAGGCTTAAGTGTAGCAACCAATCTGCTACAGCTCTCTATTAACCTTAAAT	5880
Qy	5881	GACGGAATGAGCCCTCCACCTCTATTTTGTGTGTTTGTGACAGAGCTCGGAAAGTGA	5940





Db 1021 TTAAATGACCCGAGCCTCAGGAGTAAGGTTACAAGGCTACTATTATTGGTGAATATC 1080  
QY 1081 ACTTCAATGACTTTTGAAGAGATCTGCAATGATCGATTTTGTAGAAAGATTTGAAACA 1140  
Db 1081 ACTTCAATGACTTTTGAAGAGATCTGCAATGATCGATTTTGTAGAAAGATTTGAAACA 1140  
QY 1141 ATCTGGAAGAGAGAAATGGGTGACACCTTAAGGCTGTGAATATCGGTGTGCTGCTA 1200  
Db 1141 ATCTGGAAGAGAGAAATGGGTGACACCTTAAGGCTGTGAATATCGGTGTGCTGCTA 1200  
QY 1201 AAGCAAAAGAAATGATGAGAGGTTAACAACCATCCGAGAGAGCTCCCTTGGCTTTTA 1260  
Db 1201 AAGCAAAAGAAATGATGAGAGGTTAACAACCATCCGAGAGAGCTCCCTTGGCTTTTA 1260  
QY 1261 TCTTACTTGGAGGCTCTGAGAAAGGATTTGGAATCTTTGTGACAGTGTAGATTCAGTA 1320  
Db 1261 TCTTACTTGGAGGCTCTGAGAAAGGATTTGGAATCTTTGTGACAGTGTAGATTCAGTA 1320  
QY 1321 GCAAAAGCACTGAAGCAGGCTTGAACGGGGGATCAGATATTAGAAGTAATGGCCAA 1380  
Db 1321 GCAAAAGCACTGAAGCAGGCTTGAACGGGGGATCAGATATTAGAAGTAATGGCCAA 1380  
QY 1381 ACTTGAAGAACTTTCAGCTGTCAAAAGCTATGGAATTTCTAGAATAACACACTTTAT 1440  
Db 1381 ACTTGAAGAACTTTCAGCTGTCAAAAGCTATGGAATTTCTAGAATAACACACTTTAT 1440  
QY 1441 CTATCACTGTGAACCAATTTATTGTATTAAAGACTTCTAACAGATTCTCAGAAC 1500  
Db 1441 CTATCACTGTGAACCAATTTATTGTATTAAAGAACTTCTAACAGATTCTCAGAAC 1500  
QY 1501 AGAAAAGAAATGGTGGCCCCCAGCTTCTAAATTTGGTGCATTTAAAGAGCCAGTGGCT 1560  
Db 1501 AGAAAAGAAATGGTGGCCCCCAGCTTCTAAATTTGGTGCATTTAAAGAGCCAGTGGCT 1560  
QY 1561 ACTCCATTCAGATCTTGTGTAGATGTAGAACAGGTGATAGGACTTGAAAGAGTGAACA 1620  
Db 1561 ACTCCATTCAGATCTTGTGTAGATGTAGAACAGGTGATAGGACTTGAAAGAGTGAACA 1620  
QY 1621 AAAAAAGTAAAGCCACACACTGTGGAGGAAGAACAGCTGAAAMAGTACTCCACAGAGA 1680  
Db 1621 AAAAAAGTAAAGCCACACACTGTGGAGGAAGAACAGCTGAAAMAGTACTCCACAGAGA 1680  
QY 1681 CTGGATCAGATATCTTGGCCAGAACCATATCAATGATATTGGGATTTGGACGTCTCAG 1740  
Db 1681 CTGGATCAGATATCTTGGCCAGAACCATATCAATGATATTGGGATTTGGACGTCTCAG 1740  
QY 1741 ATGACAGCATAGTAGATTTAAGGACAGAACGACATCCCACTGCAATGGCTCTCAGTG 1800  
Db 1741 ATGACAGCATAGTAGATTTAAGGACAGAACGACATCCCACTGCAATGGCTCTCAGTG 1800  
QY 1801 GAACCTTATCATCAAGTAATCTGATTTATTCAGATCACAATCGCATGCAATTTTAACTTCA 1860  
Db 1801 GAACCTTATCATCAAGTAATCTGATTTATTCAGATCACAATCGCATGCAATTTTAACTTCA 1860  
QY 1861 GTGCTACTCCTGACTTGGCAGATCAAGTCTAAGGGTTTTAAGGCTGATACGAAAGCC 1920  
Db 1861 GTGCTACTCCTGACTTGGCAGATCAAGTCTAAGGGTTTTAAGGCTGATACGAAAGCC 1920  
QY 1921 GCTACATCATGATCAGTAGAGACACTACAGCAAGGAAGTGTGATTCAGGCTATCAGGG 1980  
Db 1921 GCTACATCATGATCAGTAGAGACACTACAGCAAGGAAGTGTGATTCAGGCTATCAGGG 1980  
QY 1981 AGTTTGTCTTACAGCCCGGATCAATATTCATATGAGAGTCTCTTCACACCTG 2040  
Db 1981 AGTTTGTCTTACAGCCCGGATCAATATTCATATGAGAGTCTCTTCACACCTG 2040  
QY 2041 AGGAGTAATCAACAAAGAGACTTCCAGATCAGCTTTCCAACTTGCAGACAGAAATAC 2100  
Db 2041 AGGAGTAATCAACAAAGAGACTTCCAGATCAGCTTTCCAACTTGCAGACAGAAATAC 2100  
QY 2101 AACTGAGTGAAGGATATTATCTGAAGAAACAAACATGGAACAGAACTCTTGTTCAGATG 2160  
Db 2101 AACTGAGTGAAGGATATTATCTGAAGAAACAAACATGGAACAGAACTCTTGTTCAGATG 2160

QY 2161 AAGATGCTCAGAGAGTGTGTGAGAGAGAGTCAATTTTCCCTTCCTTCAGGTCAGACTGTGG 2220  
Db 2161 AAGATGCTCAGAGAGTGTGTGAGAGAGAGTCAATTTTCCCTTCCTTCAGGTCAGACTGTGG 2220  
QY 2221 AAGTTCGAACACAGCTCTCTATGCGAAATTTTGAACCTTTTCGCAACATTTGAACCTACTG 2280  
Db 2221 AAGTTCGAACACAGCTCTCTATGCGAAATTTTGAACCTTTTCGCAACATTTGAACCTACTG 2280  
QY 2281 AATATATGATGATTTATTTAACTCAGATCAAAAACCAAGCTGTGCCAATGGAAGAT 2340  
Db 2281 AATATATGATGATTTATTTAACTCAGATCAAAAACCAAGCTGTGCCAATGGAAGAT 2340  
QY 2341 TTGAAGAGTCTTTAACAGGAACATTTTGGGTGACATCTGAAATTTCTCAGAGAAACA 2400  
Db 2341 TTGAAGAGTCTTTAACAGGAACATTTTGGGTGACATCTGAAATTTCTCAGAGAAACA 2400  
QY 2401 ACCAGCTGAAGAGATGAAGATCATTAAGCATTTATCATAGATAGCACTGCATGTAGGG 2460  
Db 2401 ACCAGCTGAAGAGATGAAGATCATTAAGCATTTATCATAGATAGCACTGCATGTAGGG 2460  
QY 2461 AATGCAAGATTTTAACTCATATGTTTGCATATCATAGTGGCTTAAACCTGGCACAGTGG 2520  
Db 2461 AATGCAAGATTTTAACTCATATGTTTGCATATCATAGTGGCTTAAACCTGGCACAGTGG 2520  
QY 2521 CAAGACTGGGAAGCCTGGAGAAACCTCCCATTAATACGAAACATATTTCAAGATC 2580  
Db 2521 CAAGACTGGGAAGCCTGGAGAAACCTCCCATTAATACGAAACATATTTCAAGATC 2580  
QY 2581 TCCAGACCTGTTTGAATCTTCCAGAAACATGCGAATAATATGTTATGTTCTCAATAGTC 2640  
Db 2581 TCCAGACCTGTTTGAATCTTCCAGAAACATGCGAATAATATGTTATGTTCTCAATAGTC 2640  
QY 2641 AAAATATCAACCTCCCATTAATCCCTCTATTCAGATTTCAGAAAGAGATCTCACCTTCC 2700  
Db 2641 AAAATATCAACCTCCCATTAATCCCTCTATTCAGATTTCAGAAAGAGATCTCACCTTCC 2700  
QY 2701 TTACAGAGGAATGACTCAAAAGTAGAGGGCTGTCAATTTTGAAGACTAAGATGA 2760  
Db 2701 TTACAGAGGAATGACTCAAAAGTAGAGGGCTGTCAATTTTGAAGACTAAGATGA 2760  
QY 2761 TTGCAAAAGAAATTCGTACAGTGGCCGAATGGCTTCAAGTGAACATGACCTGTCCCTCA 2820  
Db 2761 TTGCAAAAGAAATTCGTACAGTGGCCGAATGGCTTCAAGTGAACATGACCTGTCCCTCA 2820  
QY 2821 TGTTCAGGACTCGGAAGAAAGAAATGGCGAGTGTGGGCTCTCAGCCAGGGTATGACAA 2880  
Db 2821 TGTTCAGGACTCGGAAGAAAGAAATGGCGAGTGTGGGCTCTCAGCCAGGGTATGACAA 2880  
QY 2881 ATGCAACAGTGTAGATGTTTGTCTCAGACAGGTGTCAATAAAGCGGGTACGTAGTT 2940  
Db 2881 ATGCAACAGTGTAGATGTTTGTCTCAGACAGGTGTCAATAAAGCGGGTACGTAGTT 2940  
QY 2941 CCTTTTCTCAATGCCAAAAGCTTTATGMAAGTCCCAANTGGCTGCAAAAGTGAACACT 3000  
Db 2941 CCTTTTCTCAATGCCAAAAGCTTTATGMAAGTCCCAANTGGCTGCAAAAGTGAACACT 3000  
QY 3001 ACCTTTCAATTTTGAAGCTGGAATGAGAGAGAGTCTTCAGACATTATCTCTCAGT 3060  
Db 3001 ACCTTTCAATTTTGAAGCTGGAATGAGAGAGAGTCTTCAGACATTATCTCTCAGT 3060  
QY 3061 GTGAGCCAGCAACCAACATTTGCTTAAGATCTTGGTGCACAAAAGCCTGTCAATTCGG 3120  
Db 3061 GTGAGCCAGCAACCAACATTTGCTTAAGATCTTGGTGCACAAAAGCCTGTCAATTCGG 3120  
QY 3121 AGACCTTCCAGTAGCTTCCAAAGGCGAGGTCACAACAAGAAAGTCAAGTCCCTGCCACAG 3180  
Db 3121 AGACCTTCCAGTAGCTTCCAAAGGCGAGGTCACAACAAGAAAGTCAAGTCCCTGCCACAG 3180  
QY 3181 CCCAGCAGCAGCCACACGACATTAATCAACAGAGGATCAAGTTCGCCGCCGTGT 3240  
Db 3181 CCCAGCAGCAGCCACACGACATTAATCAACAGAGGATCAAGTTCGCCGCCGTGT 3240

Oy 3241 CCCTTATCTTCACGGAAGAAAGTGCCTTAAGAGATCTCCACCTTTGGCATTAAC 3300  
|||||  
Db 3241 CCCTTATCTTCACGGAAGAAAGTGCCTTAAGAGATCTCCACCTTTGGCATTAAC 3300  
Oy 3301 CTCACAGCTTTAAAAAATTTCTTTCTTGTCTGAAGAGAAAGTTTGAAGCTCACA 3360  
|||||  
Db 3301 CTCACAGCTTTAAAAAATTTCTTTCTTGTCTGAAGAGAAAGTTTGAAGCTCACA 3360  
Oy 3361 AGAAGAGCTGAAGATCAATATCAATGCACTTTCAGATTTCTCTCTCTACT 3420  
|||||  
Db 3361 AGAAGAGCTGAAGATCAATATCAATGCACTTTCAGATTTCTCTCTCTACT 3420  
Oy 3421 CTCACAGATTTCTCAAGAAAGGCTATCTTGGCTCCAGTGGTACTGTGATTAAT 3480  
|||||  
Db 3421 CTCACAGATTTCTCAAGAAAGGCTATCTTGGCTCCAGTGGTACTGTGATTAAT 3480  
Oy 3481 TTTTCAGATTTCTGTACAGTGAATTTCTTACGATCCAGTAATTTAGCAATTCGTCTT 3540  
|||||  
Db 3481 TTTTCAGATTTCTGTACAGTGAATTTCTTACGATCCAGTAATTTAGCAATTCGTCTT 3540  
Oy 3541 TTGACTCAGTGCAGTCTCAGTGCAGATGAGAGGCGCAGAGGCTTCTGTACAGATCG 3600  
|||||  
Db 3541 TTGACTCAGTGCAGTCTCAGTGCAGATGAGAGGCGCAGAGGCTTCTGTACAGATCG 3600  
Oy 3601 TCGAAACAAACCTAGAGATGGGACAGATGAGAGGCGGACATGATGAACTGATCAGT 3660  
|||||  
Db 3601 TCGAAACAAACCTAGAGATGGGACAGATGAGAGGCGGACATGATGAACTGATCAGT 3660  
Oy 3661 ATAGCTTGGGGTCTTATGACCAATGTCGAGGGCGGAGGCTTATATGCTTACAGTACAG 3720  
|||||  
Db 3661 ATAGCTTGGGGTCTTATGACCAATGTCGAGGGCGGAGGCTTATATGCTTACAGTACAG 3720  
Oy 3721 TAAATTTCTTCTCAAGCAGACAGAGAACTTTCCAGAGATCAGGGGATGGCGGTCACTTG 3780  
|||||  
Db 3721 TAAATTTCTTCTCAAGCAGACAGAGAACTTTCCAGAGATCAGGGGATGGCGGTCACTTG 3780  
Oy 3781 ATGCTGCTGACAGTGGCGGTGGGAGCTGAGCTCATGCTCAAGTGGTCCCATGATTAATA 3840  
|||||  
Db 3781 ATGCTGCTGACAGTGGCGGTGGGAGCTGAGCTCATGCTCAAGTGGTCCCATGATTAATA 3840  
Oy 3841 TACAGACGATCCAGACACAGAGAACTGGGAGACTTTCCATTCGGGATACTCACTTTG 3900  
|||||  
Db 3841 TACAGACGATCCAGACACAGAGAACTGGGAGACTTTCCATTCGGGATACTCACTTTG 3900  
Oy 3901 ATTATTCAGGGGATCCTGAGGTTTATGGGATCAAGACGCAATGAGCCAAATTAATGT 3960  
|||||  
Db 3901 ATTATTCAGGGGATCCTGAGGTTTATGGGATCAAGACGCAATGAGCCAAATTAATGT 3960  
Oy 3961 TTTCTGATCATAGCAAAAGTATTAACAGGCAAAATCAAAAGTAGAGAGAGCTTTGAACAAG 4020  
|||||  
Db 3961 TTTCTGATCATAGCAAAAGTATTAACAGGCAAAATCAAAAGTAGAGAGAGCTTTGAACAAG 4020  
Oy 4021 CCCAATCCCCGAGCAGAGCTGGGCGTCTTCCACAGGTTACTGGGGAAGAGCTCAGAAAGTG 4080  
|||||  
Db 4021 CCCAATCCCCGAGCAGAGCTGGGCGTCTTCCACAGGTTACTGGGGAAGAGCTCAGAAAGTG 4080  
Oy 4081 ACACAGGCAATTAAGGAGGAGGAGGAGGATGTTTCCATTGAAGCCGAAGACAGTA 4140  
|||||  
Db 4081 ACACAGGCAATTAAGGAGGAGGAGGAGGATGTTTCCATTGAAGCCGAAGACAGTA 4140  
Oy 4141 GCCTAAGCTGTGTGACTAGGAAGAACCAAGCCTGTCCCATGCTGCCCATATAGCTG 4200  
|||||  
Db 4141 GCCTAAGCTGTGTGACTAGGAAGAACCAAGCCTGTCCCATGCTGCCCATATAGCTG 4200  
Oy 4201 TGGCATCAAGTACTCAAAAGGGGCTCATTTGCAGAAAGAGGGGAGGATTCAGAGAGCCC 4260  
|||||  
Db 4201 TGGCATCAAGTACTCAAAAGGGGCTCATTTGCAGAAAGAGGGGAGGATTCAGAGAGCCC 4260  
Oy 4261 CGCCACCCCTCCGGGCTACATTTGGAATTCCTCATTTGATGAGGAGGAGGAGGAGGAGG 4320  
|||||  
Db 4261 CGCCACCCCTCCGGGCTACATTTGGAATTCCTCATTTGATGAGGAGGAGGAGGAGGAGG 4320  
Oy 4321 ATTCAGCCAGGAAGACGCGGAGCTACAAAGGTGGCCCTTCAGAGATCGCGGATGTGCGAC 4380  
|||||

Db 4321 ATTCAGCCAGGAAGACGCGGAGCTACAAAGGTGGCCCTTCAGAGATCGCGGATGTGCGAC 4380  
Oy 4381 GATCTCCGACACACACTGGGCTTATCGGTACAGACAGCCACATGGGCAATCCACACAGA 4440  
|||||  
Db 4381 GATCTCCGACACACACTGGGCTTATCGGTACAGACAGCCACATGGGCAATCCACACAGA 4440  
Oy 4441 GCAGGCTGTGAACAACCTCAGTGGCATTAACCGAAGAGCTGACCGGCGCTGCGCC 4500  
|||||  
Db 4441 GCAGGCTGTGAACAACCTCAGTGGCATTAACCGAAGAGCTGACCGGCGCTGCGCC 4500  
Oy 4501 CTATCAGTCCCAAGGTTTTCACACGAGAGATGAAGATGAACAAGTTTCTGCTGTTT 4560  
|||||  
Db 4501 CTATCAGTCCCAAGGTTTTCACACGAGAGATGAAGATGAACAAGTTTCTGCTGTTT 4560  
Oy 4561 GAGGACAGACTTTTCTGGAAGCAGAGAGGAGCCACCTGAAGAGAGACACAAAGAGCT 4620  
|||||  
Db 4561 GAGGACAGACTTTTCTGGAACAGAGAGGAGCCACCTGAAGAGAGCACAAGAGAGCT 4620  
Oy 4621 CCTGAGCATTTGAGGCTTGGAACTCACAATTCAGAGAGGAGGAGGAGGAGGAGGAGGAGG 4680  
|||||  
Db 4621 CCTGAGCATTTGAGGCTTGGAACTCACAATTCAGAGAGGAGGAGGAGGAGGAGGAGGAGG 4680  
Oy 4681 CCTGCTTAAAAAGCAGCATGGGGCTTTCTTCCCTTCCCTTCCCTTTCATGTGA 4740  
|||||  
Db 4681 CCTGCTTAAAAAGCAGCATGGGGCTTTCTTCCCTTCCCTTCCCTTTCATGTGA 4740  
Oy 4741 AATACTGTGAAGAAATTTGCCCTGGCAGCTTTTCAGACTTTGTTGCTTGAATGCACAGTGC 4800  
|||||  
Db 4741 AATACTGTGAAGAAATTTGCCCTGGCAGCTTTTCAGACTTTGTTGCTTGAATGCACAGTGC 4800  
Oy 4801 ACCAATCTTCGAGCTCCACAGTGTGCTGCTGCACATACAGATATTCGCAATTC 4860  
|||||  
Db 4801 ACCAATCTTCGAGCTCCACAGTGTGCTGCTGCACATACAGATATTCGCAATTC 4860  
Oy 4861 CAAGATCATCACACAGAGATGATCACTCTGAGCTGACATTCGCAATTCGCTGGAAGATTT 4920  
|||||  
Db 4861 CAAGATCATCACACAGAGATGATCACTCTGAGCTGACATTCGCAATTCGCTGGAAGATTT 4920  
Oy 4921 TTTTAAATCTCTTTTATGATTTCAATCCAGTCTGACAGCTTTGATGATGATG 4980  
|||||  
Db 4921 TTTTAAATCTCTTTTATGATTTCAATCCAGTCTGACAGCTTTGATGATGATG 4980  
Oy 4981 AGAAAGCTAGGCAATGAACTCTTGGGGGCTTTAAACCCACAGGAAGACAAAGAAAA 5040  
|||||  
Db 4981 AGAAAGCTAGGCAATGAACTCTTGGGGGCTTTAAACCCACAGGAAGACAAAGAAAA 5040  
Oy 5041 CAATGAATCTTTTGAAGTACAGTGTGCTGCACTTGTATACATGCTCTTTAAAAA 5100  
|||||  
Db 5041 CAATGAATCTTTTGAAGTACAGTGTGCTGCACTTGTATACATGCTCTTTAAAAA 5100  
Oy 5101 AAAAAATGAGTTTAAAGATTTTGTTCAGAGAGTAAATATATATATATATATATATATATAT 5160  
|||||  
Db 5101 AAAAAATGAGTTTAAAGATTTTGTTCAGAGAGTAAATATATATATATATATATATATATAT 5160  
Oy 5161 TATTAATTTTAAACCTTAAGTAGGTTGACAGGCTGTTTCTGAAGAACCAAAATATGCGGG 5220  
|||||  
Db 5161 TATTAATTTTAAACCTTAAGTAGGTTGACAGGCTGTTTCTGAAGAACCAAAATATGCGGG 5220  
Oy 5221 ACAGGGTGTGGCAGACCAAGAAAGCGGGAAGAGCTGTGACCTTGAGCTTGCCATG 5280  
|||||  
Db 5221 ACAGGGTGTGGCAGACCAAGAAAGCGGGAAGAGCTGTGACCTTGAGCTTGCCATG 5280  
Oy 5281 TCCCTTCTGTCTCACCAGGAGAGTGCCTATCTCTGAAGATGAAGATTTTACCAATTA 5340  
|||||  
Db 5281 TCCCTTCTGTCTCACCAGGAGAGTGCCTATCTCTGAAGATGAAGATTTTACCAATTA 5340  
Oy 5341 TACCAAGACCTCATCTGCTCTTCCAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5400  
|||||  
Db 5341 TACCAAGACCTCATCTGCTCTTCCAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5400  
Oy 5401 CACATGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5460  
|||||

Db 5401 CACATGCGCCAGGGGAGGAGAACTAGACACCTTGTGTCCTGAGCCCTTATGAGGCGAG 5460  
Qy 5461 ACGGTGTCATTTGGCGGATGTCCTGTCCTCATTTGAGATGGATGGCAACCCCATTTTAA 5520  
Db 5461 ACGGTGTCATTTGGCGGATGTCCTGTCCTCATTTGAGATGGATGGCAACCCCATTTTAA 5520  
Qy 5521 GTTATATTTCTTTGATTTTGTAAATTTAGAGGTGATGTTTGTGTTTGTGTTTGTGTT 5580  
Db 5521 GTTATATTTCTTTGATTTTGTAAATTTAGAGGTGATGTTTGTGTTTGTGTTTGTGTT 5580  
Qy 5581 TTTTAAAGAGAAACATTTATTAACAGTGGATGAGTGGAGTGAACAGCAGCTTGGATGTT 5640  
Db 5581 TTTTAAAGAGAAACATTTATTAACAGTGGATGAGTGGAGTGAACAGCAGCTTGGATGTT 5640  
Qy 5641 GGAGCTAATGCCAGCTGTTTATCTGCTCTTTCCAGACAGCCCTCCTTTATTTGAATTGGC 5700  
Db 5641 GGAGCTAATGCCAGCTGTTTATCTGCTCTTTCCAGACAGCCCTCCTTTATTTGAATTGGC 5700  
Qy 5701 ATTAGGGAATTAACAAGCCCTTAAAGCTGATTAAGAATCAAAAACCTGGTTACATGGC 5760  
Db 5701 ATTAGGGAATTAACAAGCCCTTAAAGCTGATTAAGAATCAAAAACCTGGTTAGACATGCC 5760  
Qy 5761 AGCCTTTCGAAGCAGGTAGTGTACCAAAAGCTAACCTCCAAAGTGGCTTTATGAGCGCTG 5820  
Db 5761 AGCCTTTCGAAGCAGGTAGTGTACCAAAAGCTAACCTCCAAAGTGGCTTTATGAGCGCTG 5820  
Qy 5821 CATATAGAGAAGGCTTAAGTGTAGCAACATCTGCTCAGCAGCTGCTATTAACCTTATAT 5880  
Db 5821 CATATAGAGAAGGCTTAAGTGTAGCAACATCTGCTCAGCAGCTGCTATTAACCTTATAT 5880  
Qy 5881 GACGTGAATAACCCCTGCTGCTGATTTTGTGTTGTTTGCACAGAGCTCCGGAAAAGTGA 5940  
Db 5881 GACGTGAATAACCCCTGCTGCTGATTTTGTGTTGTTTGCACAGAGCTCCGGAAAAGTGA 5940  
Qy 5941 AGCGTGCCTAATCTGAGTACTCAATCTGAGGAACTGCTGCTGTTGATTTTGTTC 6000  
Db 5941 AGCGTGCCTAATCTGAGTACTCAATCTGAGGAACTGCTGCTGCTGTTGATTTTGTTC 6000  
Qy 6001 ATTTAATTCAGCTGATCATATGATCAGTATGATTAAGCTTAATAGCTCAAAATTTTAAA 6060  
Db 6001 ATTTAATTCAGCTGATCATATGATCAGTATGATTAAGCTTAATAGCTCAAAATTTTAAA 6060  
Qy 6061 GTGGAATTCAGTGTGTTTTCACGTATCAAAAGTGTCAAGTGTGCTTTATTTAATTTCT 6120  
Db 6061 GTGGAATTCAGTGTGTTTTCACGTATCAAAAGTGTCAAGTGTGCTTTATTTAATTTCT 6120  
Qy 6121 CTCTGTATCATGSCATTTGCTACTGCTTATTAACATTTGCAATTTATGCAATTTGTAATT 6180  
Db 6121 CTCTGTATCATGSCATTTGCTACTGCTTATTAACATTTGCAATTTATGCAATTTGTAATT 6180  
Qy 6181 TTACATGTAATATGCAATTTATTTGTCAGTATTAATATAGGCTATGAGCCATGTGCAT 6240  
Db 6181 TTACATGTAATATGCAATTTATTTGTCAGTATTAATATAGGCTATGAGCCATGTGCAT 6240  
Qy 6241 ATAGAAAGACAGAAATCTAGCTCTACGACAAAGTGTGCAAAATGTTATCTAAGCAATTAAGT 6300  
Db 6241 ATAGAAAGACAGAAATCTAGCTCTACGACAAAGTGTGCAAAATGTTATCTAAGCAATTAAGT 6300  
Qy 6301 AATGTAGACATAGAGCTGCTAATCTCAGTTCGCTGCTGATGTCAAGTGCAGAAATGTA 6360  
Db 6301 AATGTAGACATAGAGCTGCTAATCTCAGTTCGCTGCTGATGTCAAGTGCAGAAATGTA 6360  
Qy 6361 CAATTAAGTGTGATTTCTCATATCTTTGATGATGATGATGATGATGATGATGATGATGAT 6420  
Db 6361 CAATTAAGTGTGATTTCTCATATCTTTTGTATGATGATGATGATGATGATGATGATGAT 6420  
Qy 6421 AGACATTTGGTGGAGTCTGATCCCTTTGATTTTAAATTAACAAATTAATTTGATATTTGGT 6480  
Db 6421 AGACATTTGGTGGAGTCTGATCCCTTTGATTTTAAATTAACAAATTAATTTGATATTTGGT 6480  
Qy 6481 TATATTTTGTGTTGAAGATGATGAATATGATATGATGATGATGATGATGATGATGATGAT 6540  
Db 6481 TATATTTTGTGTTGAAGATGATGAATATGATATGATGATGATGATGATGATGATGATGAT 6540

Qy 6541 CACCTGCAAAATTAATTAATTAACAT 6568  
Db 6541 CACCTGCAAAATTAATTAATTAACAT 6568  
RESULT 3  
AB002311 6568 bp mRNA linear PRI 06-OCT-2001  
LOCUS  
DEFINITION Human mRNA for KIAA0313 gene, complete cds.  
ACCESSION AB002311  
VERSION AB002311.1 GI:2224566  
KEYWORDS  
SOURCE Homo sapiens male brain cDNA to mRNA, clone lib:pbluescriptII SK plus clone:H0186.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Mammalia; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
1 Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O. Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro  
DNA Res. 4 (2), 141-150 (1997)  
97349984  
JOURNAL MEDLINE  
REFERENCE  
AUTHORS Ohara,O., Nagase,T., Kikuno,R. and Nomura,N.  
TITLE Direct Submission  
SUBMITTED (28-MAR-1997) Osamu Ohara, Kazusa DNA Research Institute; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:cdna@foc.kazusa.or.jp, Tel:+81-438-52-3913)  
JOURNAL  
FEATURES  
source  
1, 6568  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="H0186"  
/sex="male"  
/tissue="brain"  
/clone.lib="pbluescriptII SK plus"  
1, 6568  
/gene="KIAA0313"  
63, .4562  
/codon.start=1  
/protein.id="BAA20772.1"  
/db\_xref="GI:2224567"  
/translation="MKPLAIPANHGVMGOEKHSLPADFTKLHLTDLSPQVTHVSSS HSGCSTSSGSSLSIDYQATESPAGMDLSGLPETAVIDSEDDDEDEIRASDPLM SMDIVDCLEKDPIDRTDDTLEQLPEPHQLPAPRNMYSVRELCAVMYFAVYBRG TIVLDGEBLDSVILNGSVEVITPDKAEILTCNGNSFVSP7MDKRYTKGWRTKY DDCQFVCIADQDYCRILNOVEKRMQKVEEGEIVWKRHRLDRTRGRGHIIVKGTIS ERLTHLVHEHSDVDPTEIEDLTVRFELSPMEVGKLLLEWNPDLRKVTVVYL LVMNHPNDFEGDPAMTRLEFEENLRERKMGHLRLINJACAKARLMRTLPKPS REAPDPTILGSEKGFQFVNDVSNGSKATGKLRGDDILLEVNGAFENIOLSKAM ELRLNTHLSTIVKTNLVEFKELRLSEBRKNQAPHLPKIGDIKKARSYSLPDAL VQOVGLEKVNKRSKANTVGGKRNKLTLDRTRISILQKRPNDIGIOSDDSLVGL ROTKHIPALPVSGTSSNBDLQSHRITLDFSA7PDLPPQVLEVFADDOSTRIMT SKDTAKEVYIOAIRFEATATAPDQSLCEVSPEVGIQKRLPDDQLSLADRIQLS SKRYLYKNMNETLCSDEDAELLRESOISLQSLSTVAVATOLSKRNELERINPEPE YIDDLFKLSKTSKTSICNLAPVARKLTTEMEKLPNKYEKLPQDQDLFDSRMARKRVN RECKNFSMFALISGILNAPVARKLTTEMEKLPNKYEKLPQDQDLFDSRMARKRVN LNSONLQPIILFLPVIKDLTLEHGNDKVDGLVNEKRLMKIKETRIHYGRASVN MPDLMALFRTKKRKRSLGSLQSGSTNATVLDVAQGTGKRRKRVRSFLINAKRLVDAQ MARKVOYLSNLELWDEESLQTLQCEPATNLPKNKGRKPKVSETPASVAPRQIS OOKAOSLPPQPOOPPAHKLINGLOVAPVSLSPKRVKVPKDLPPFGINSPOALKIL SLSEBSLLERRHKOAEPTISNLSQLSPPSPSSPKRGYTLAPSGTVDNDSGHS ELSSRSIVSNSSFDSVPVSLDERQRHRSVITVTNMGKMERMTIPEDOYSIGS IAPMSGRKLIVATVIVISSPTEELSDQGDASIDADSGKSTSCSSGSHDIOT IAHORSWETLPGHTHFDYSDPAGLMASSHMDIMFSHSTKYNRQNSHESLEQA OSRAWSSSTGYWGEDSEGDGTIKRGGKDVSIIESSHLSVTTEBTKVPVPAHRI AVASSWTKLIARKEGRYREPPTPGYIGIPTDEPESHSHARKPPDYVALQORIR MARSSDTAGPSSVQOPHGHPTSSRVNKPQWHRKNESDPRLAPOSGGFEDEDE



Db 2041 AGGAGTAAATCAACAAGAGACTTCAGATTCAGCTTTCACAACTTGGAGACAGATAC 2100  
Qy 2101 AACTGAGTGGAGATATATCTGAAAAACACATGAAAAACAGAACTCTTGTTCAGATG 2160  
Db 2101 AACTGAGTGGAGATATATCTGAAAAACACATGAAAAACAGAACTCTTGTTCAGATG 2160  
Qy 2161 AGATGCTCAGAGGTTGTGAGAGAGATCAATTTCCCTCTTCAGCTCAGACAGTGG 2220  
Db 2161 AGATGCTCAGAGGTTGTGAGAGAGATCAATTTCCCTCTTCAGCTCAGACAGTGG 2220  
Qy 2221 AAGTTGCAACACAGCTCTCTATGCGAAATTTTGAACCTTTTGCACATTTGAACCTACTG 2280  
Db 2221 AAGTTGCAACACAGCTCTCTATGCGAAATTTTGAACCTTTTGCACATTTGAACCTACTG 2280  
Qy 2281 AATATATAGATGATTTATTTTAACTCAGATCAAAAACACAGCTGCGAACCTTAAGAT 2340  
Db 2281 AATATATAGATGATTTATTTTAACTCAGATCAAAAACACAGCTGCGAACCTTAAGAT 2340  
Qy 2341 TTGAGAGATCAATTAACAGAGAAATTTGGGAGCATGGAATTTCTCAGAGAAACA 2400  
Db 2341 TTGAGAGATCAATTAACAGAGAAATTTGGGAGCATGGAATTTCTCAGAGAAACA 2400  
Qy 2401 ACCAGCTGAGAGAGATGAGATCATTAAGCATTTTCATCAAGATAGCATGCTGATAGG 2460  
Db 2401 ACCAGCTGAGAGAGATGAGATCATTAAGCATTTTCATCAAGATAGCATGCTGATAGG 2460  
Qy 2461 AATGCAAGATTTTAACTCAATGTTTGCATCATGAGCTGACCTTAACCTGACACAGTGG 2520  
Db 2461 AATGCAAGATTTTAACTCAATGTTTGCATCATGAGCTGACCTTAACCTGACACAGTGG 2520  
Qy 2521 CAAGACTGCGAAGCCTGGAGAACTTCCCAATTAATACGAAAACTATTTTCANAGATC 2580  
Db 2521 CAAGACTGCGAAGCCTGGAGAACTTCCCAATTAATACGAAAACTATTTTCANAGATC 2580  
Qy 2581 TCCAAAGACTGTTTATCTTCCAGAAACATGSCAAAAATTCGTAATGTTCTCAATAGTC 2640  
Db 2581 TCCAAAGACTGTTTATCTTCCAGAAACATGSCAAAAATTCGTAATGTTCTCAATAGTC 2640  
Qy 2641 AAAATCTACAACTCCCATATATCCCTCTATTTCCAGTTATCAAAAAGATCTCACCTTC 2700  
Db 2641 AAAATCTACAACTCCCATATATCCCTCTATTTCCAGTTATCAAAAAGATCTCACCTTC 2700  
Qy 2701 TTCAAGAAAGATGACTCAAAAAGTACGAGGCTGTCATTTTGAAGAGTAAAGATGA 2760  
Db 2701 TTCAAGAAAGATGACTCAAAAAGTACGAGGCTGTCATTTTGAAGAGTAAAGATGA 2760  
Qy 2761 TTGCAAAAAGATTCGTACGCTGGCGAATGCTTCAAGTACATGAGACCTCCCTCA 2820  
Db 2761 TTGCAAAAAGATTCGTACGCTGGCGAATGCTTCAAGTACATGAGACCTCCCTCA 2820  
Qy 2821 TGTTCAGAGACTCGAAGAAATGGCGAGTTTGGGGTCTCTAGCCAGGGTAGTACAA 2880  
Db 2821 TGTTCAGAGACTCGAAGAAATGGCGAGTTTGGGGTCTCTAGCCAGGGTAGTACAA 2880  
Qy 2881 ATGCAACAGTCTGATGATGTTGCTGACAGAGTGTGATTAAGAAACGGGGTAGTGTATG 2940  
Db 2881 ATGCAACAGTCTGATGATGTTGCTGACAGAGTGTGATTAAGAAACGGGGTAGTGTATG 2940  
Qy 2941 CCTTTCTCAATGCGAAAAAGCTTTATGAAGATGCGCAATGGCTCGAAAAATGGAAGCAGT 3000  
Db 2941 CCTTTCTCAATGCGAAAAAGCTTTATGAAGATGCGCAATGGCTCGAAAAATGGAAGCAGT 3000  
Qy 3001 ACCTTTCAATTTGGAGTAGAATGAGAGAGAGTCTTCAGACATTTATCTCTGAGT 3060  
Db 3001 ACCTTTCAATTTGGAGTAGAATGAGAGAGAGTCTTCAGACATTTATCTCTGAGT 3060  
Qy 3061 GTGAGCCAGCAACAACATTTGCTTAAGAAATCTGTGACAAAAAGCCTGTCAATTCG 3120  
Db 3061 GTGAGCCAGCAACAACATTTGCTTAAGAAATCTGTGACAAAAAGCCTGTCAATTCG 3120  
Qy 3121 AGACCTCTCAGTACGTCAGAGGAGGCTCACACAGAAAGCTCAGTCCCTGCCACAGC 3180  
Db 3121 AGACCTCTCAGTACGTCAGAGGAGGCTCACACAGAAAGCTCAGTCCCTGCCACAGC 3180

Qy 3181 CCCAGAGAGCCACACAGACATTAATATCAACAGGACTACAGGTTCCCGCGTGT 3240  
Db 3181 CCCAGAGAGCCACACAGACATTAATATCAACAGGACTACAGGTTCCCGCGTGT 3240  
Qy 3241 CCCTTATCTCTCAGAGAAAGAGTCCGTAAGAGATCTCCACCTTTTGGCATAACT 3300  
Db 3241 CCCTTATCTCTCAGAGAAAGAGTCCGTAAGAGATCTCCACCTTTTGGCATAACT 3300  
Qy 3301 CTCACACAGCTTTAAAAAATTTCTTGTGTGGAAGAGAAAGTTTGGAGCTGCA 3360  
Db 3301 CTCACACAGCTTTAAAAAATTTCTTGTGTGGAAGAGAAAGTTTGGAGCTGCA 3360  
Qy 3361 AGAAACAGCTGAGATACATATCAATATCAATCTTCCAGCTTTCTCTCTCTACTT 3420  
Db 3361 AGAAACAGCTGAGATACATATCAATATCAATCTTCCAGCTTTCTCTCTCTACTT 3420  
Qy 3421 CTCACAGAGTTCTCCAGAGAAAGGCTATCTTGGCTCCAGTGTACTGTGATATT 3480  
Db 3421 CTCACAGAGTTCTCCAGAGAAAGGCTATCTTGGCTCCAGTGTACTGTGATATT 3480  
Qy 3481 TTTCAAGATCTGTCCAGAGTGAATTTCTTCAAGATTCAGATTTGTTAGCAATTCGCTT 3540  
Db 3481 TTTCAAGATCTGTCCAGAGTGAATTTCTTCAAGATTCAGATTTGTTAGCAATTCGCTT 3540  
Qy 3541 TTGACTCAGTGGCAGTCTCAGTACATGACAGATGAGAGGCGCCAGAGGATCTGTCAACATCG 3600  
Db 3541 TTGACTCAGTGGCAGTCTCAGTACATGACAGATGAGAGGCGCCAGAGGATCTGTCAACATCG 3600  
Qy 3601 TGGAAACAAACCTAGGAGATGGGAGATGAGAGGCGGACATGATTTGAACCTGATACGT 3660  
Db 3601 TGGAAACAAACCTAGGAGATGGGAGATGAGAGGCGGACATGATTTGAACCTGATACGT 3660  
Qy 3661 ATGAGCTGGGGTCTATGACCAATGTCCAGAGGCGGAGGCTTATATGCTACAGCTACAG 3720  
Db 3661 ATGAGCTGGGGTCTATGACCAATGTCCAGAGGCGGAGGCTTATATGCTACAGCTACAG 3720  
Qy 3721 TAAATTTCTTCTCCAGACAGAGAGAACTTTCCAGAGATGAGGGGATCGCGCTCACCTTG 3780  
Db 3721 TAAATTTCTTCTCCAGACAGAGAGAACTTTCCAGAGATGAGGGGATCGCGCTCACCTTG 3780  
Qy 3781 ATGCTGCTCAGATGGCGCTGGAGAGCTGAGAGCTCATGCTCAAGTGGCTCCCATGATATA 3840  
Db 3781 ATGCTGCTCAGATGGCGCTGGAGAGCTGAGAGCTCATGCTCAAGTGGCTCCCATGATATA 3840  
Qy 3841 TACAGAGATCCAGACACAGAGAGAGCTGAGAGACTTTCATTCGGGCACTACTCTTG 3900  
Db 3841 TACAGAGATCCAGACACAGAGAGAGCTGAGAGACTTTCATTCGGGCACTACTCTTG 3900  
Qy 3901 ATTATTCAGGGGATCTGACAGTATTATGGGATCAAGCAGCCATATGACCAATATATG 3960  
Db 3901 ATTATTCAGGGGATCTGACAGTATTATGGGATCAAGCAGCCATATGACCAATATATG 3960  
Qy 3961 TTTTCTGATCTACACAAATATATACAGGCAAAAATCAAAAGTACAGAGAGCTTGAACAAG 4020  
Db 3961 TTTTCTGATCTACACAAATATATACAGGCAAAAATCAAAAGTACAGAGAGCTTGAACAAG 4020  
Qy 4021 CCCAGTCCGAGAGAGCTGGCGCTCTTCCACAGATTACTGGGAGAAAGCTCAGAAAGTGG 4080  
Db 4021 CCCAGTCCGAGAGAGCTGGCGCTCTTCCACAGATTACTGGGAGAAAGCTCAGAAAGTGG 4080  
Qy 4081 ACACAGGACATTAAGCGGAGGAGTGAAGAGATGTTTCATTTGAAGCCGAAAGCAGTA 4140  
Db 4081 ACACAGGACATTAAGCGGAGGAGTGAAGAGATGTTTCATTTGAAGCCGAAAGCAGTA 4140  
Qy 4141 GCTTAAGCTGTGATACGAGAAACCAAGCTGTGCCATGCTCTGCCACATAGCTG 4200  
Db 4141 GCTTAAGCTGTGATACGAGAAACCAAGCTGTGCCATGCTCTGCCACATAGCTG 4200  
Qy 4201 TGGCATCAAGTACTACAAAGGGGCTCATTTGCAGAAAGGAGGGAGGTATCGAGAGCCCC 4260  
Db 4201 TGGCATCAAGTACTACAAAGGGGCTCATTTGCAGAAAGGAGGGAGGTATCGAGAGCCCC 4260



4261 CGCCACCCCTCCGGCTACATTGGAATTCCTTACGATCTTCCAGAGGCGACTCCC 4320  
4261 CGCCACCCCTCCGGCTACATTGGAATTCCTTACGATCTTCCAGAGGCGACTCCC 4320  
4321 ATCCAGCCAGGAACCAGCGGACTACACGCGGCTTTCAGAGATGCGGATGCTGCAC 4380  
4321 ATCCAGCCAGGAACCAGCGGACTACACGCGGCTTTCAGAGATGCGGATGCTGCAC 4380  
4381 GATTCCTCCAGACAGCTGGGCTTCATCCGCTACAGAGCCACATGGGCAATCCACAGCA 4440  
4381 GATTCCTCCAGACAGCTGGGCTTCATCCGCTACAGAGCCACATGGGCAATCCACAGCA 4440  
4441 GCAGGCTGTGAACAACCTCAGTGGCATTAACCGAAGAGCTGACCGCGCTCGGCC 4500  
4441 GCAGGCTGTGAACAACCTCAGTGGCATTAACCGAAGAGCTGACCGCGCTCGGCC 4500  
4501 CTTATCAGTCCCAAGGGTTTCCACGAGAGAGATGAAGATGAACAAGTTTCTGCTGTTT 4560  
4501 CTTATCAGTCCCAAGGGTTTCCACGAGAGAGATGAAGATGAACAAGTTTCTGCTGTTT 4560  
4561 GAGGCAACACTTTTCTGAGAGCAGAGCGACCTGAAAGAGAGCAACAGAGAGCT 4620  
4561 GAGGCAACACTTTTCTGAGAGCAGAGCGACCTGAAAGAGAGCAACAGAGAGCT 4620  
4621 CCTGACATTTGGAGCCTTGGAACTCATTCTGAGAGAGGTGGACCGTTCCTCTTC 4680  
4621 CCTGACATTTGGAGCCTTGGAACTCATTCTGAGAGAGGTGGACCGTTCCTCTTC 4680  
4681 CCTGCTTTAAAGACAGCATGGGGCTTCTCCCTCTTCTCTTCCCTTTCGATGATGA 4740  
4681 CCTGCTTTAAAGACAGCATGGGGCTTCTCCCTCTTCTCTTCCCTTTCGATGATGA 4740  
4741 AATACGTGAAGAATTTGCCCTGGCACTTTTACAGACTTTGTTGCTTGAATGCAAGTGC 4800  
4741 AATACGTGAAGAATTTGCCCTGGCACTTTTACAGACTTTGTTGCTTGAATGCAAGTGC 4800  
4801 AGCAATCTTGAAGCTCCACTGTTGCTGCTGCACATCAGACAGATGATTCCTCAATTC 4860  
4801 AGCAATCTTGAAGCTCCACTGTTGCTGCTGCACATCAGACAGATGATTCCTCAATTC 4860  
4861 CAGATCATCACAACAAGATGATTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4920  
4861 CAGATCATCACAACAAGATGATTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4920  
4921 TTTTATATCTCTCTTTTAAATTTCAATTCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 4980  
4921 TTTTATATCTCTCTTTTAAATTTCAATTCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 4980  
4981 AGAAAGCTAGCATTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5040  
4981 AGAAAGCTAGCATTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5040  
5041 CAATGAATCTTTGAGTACAGTGTGTCACACTGTTTACAAATGTCCTCTTTAAAAA 5100  
5041 CAATGAATCTTTGAGTACAGTGTGTCACACTGTTTACAAATGTCCTCTTTAAAAA 5100  
5101 AAAAATGAGTTTAAAGATTTGTTTCAGAGAGTAATATATATATATATATATATATATAT 5160  
5101 AAAAATGAGTTTAAAGATTTGTTTCAGAGAGTAATATATATATATATATATATATATAT 5160  
5161 TTTTATTTTAAACCTTAAGTGGTGGCAGCTGCTGTTCTGAAAAACCAATATGCGCG 5220  
5161 TTTTATTTTAAACCTTAAGTGGTGGCAGCTGCTGTTCTGAAAAACCAATATGCGCG 5220  
5221 ACAGGCTGGGACACACAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5280  
5221 ACAGGCTGGGACACACAGAAG 5280  
5281 TCGTTTGTGTACCCGCGGAAGTGGCTTATCTGGAAGATGAATGTTAGCAATTTAA 5340  
5281 TCGTTTGTGTACCCGCGGAAGTGGCTTATCTGGAAGATGAATGTTAGCAATTTAA 5340  
5341 TACGAAGACACGTATCTGCTCTTCCAGAGTGGATGGGTTCTTGTATAAACCTTTTG 5400

5341 TACGAAGACACGTATCTGCTCTTCCAGAGTGGATGGGTTCTTGTATAAACCTTTTG 5400  
5401 CACATGGCCAGGGGAGGAGACTAGACCTTGTGTCTGTCTGAGGCTTATGAGAGCAG 5460  
5401 CACATGGCCAGGGGAGGAGACTAGACCTTGTGTCTGTCTGAGGCTTATGAGAGCAG 5460  
5401 CACATGGCCAGGGGAGGAGACTAGACCTTGTGTCTGTCTGAGGCTTATGAGAGCAG 5460  
5461 ACCGTGCAATGGCGGATGTGCTCTCTCATTGAGATGAGATGGCAACCCCAATTTTAA 5520  
5461 ACCGTGCAATGGCGGATGTGCTCTCTCATTGAGATGAGATGGCAACCCCAATTTTAA 5520  
5521 GTTATATTTCTTTGATTTTGTAAATTTAGAGTGGAGTTTGTGTTTGTGTTTGTGTT 5580  
5521 GTTATATTTCTTTGATTTTGTGTTAATTTAGAGTGGAGTTTGTGTTTGTGTTTGTGTT 5580  
5581 TTTTATTTTAAAGAACATTTATATAGTATGATGATGATGATGATGATGATGATGATGAT 5640  
5581 TTTTATTTTAAAGAACATTTATATAGTATGATGATGATGATGATGATGATGATGATGAT 5640  
5641 GAGCTAATGCCAGCTGTTTATATCTGCTTTCAGACAGCTCCCTTTATGTAATTTGGC 5700  
5641 GAGCTAATGCCAGCTGTTTATATCTGCTTTCAGACAGCTCCCTTTATGTAATTTGGC 5700  
5701 ATTAGGGAATTAACAAGCCTTTAAAGTGAATAAGATCAAAAACCTGGTTAGACATGCC 5760  
5701 ATTAGGGAATTAACAAGCCTTTAAAGTGAATAAGATCAAAAACCTGGTTAGACATGCC 5760  
5761 AGCCTTTGCAAGGACAGTATAGTACCAAGAGACTCAACCTCAAGTGGCTTATGAGAGCTG 5820  
5761 AGCCTTTGCAAGGACAGTATAGTACCAAGAGACTCAACCTCAAGTGGCTTATGAGAGCTG 5820  
5821 CATATGAGAGAGCTTAAAGTATGACACCAATCTGTCACAGCTGCTATTAATACCTTAAT 5880  
5821 CATATGAGAGAGCTTAAAGTATGACACCAATCTGTCACAGCTGCTATTAATACCTTAAT 5880  
5881 GACTGAATGACCCCTCCACTATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 5940  
5881 GACTGAATGACCCCTCCACTATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 5940  
5941 AGGCTCCCAATCTGAGTACTCAATATGAGAGAACTGCTGCTGCTGCTGCTGCTGCTGCTG 6000  
5941 AGGCTCCCAATCTGAGTACTCAATATGAGAGAACTGCTGCTGCTGCTGCTGCTGCTGCTG 6000  
6001 ATTAAATCAGCTGATCATATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 6060  
6001 ATTAAATCAGCTGATCATATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 6060  
6061 GTGGAATTCAGAGTGTGTTTCACTGATATCAAAACATGCTGCTGCTGCTGCTGCTGCTGCTG 6120  
6061 GTGGAATTCAGAGTGTGTTTCACTGATATCAAAACATGCTGCTGCTGCTGCTGCTGCTGCTG 6120  
6121 CTTCTGATCATGAGCTTGTGCTACTGCTGCTTATTAATGCTCAATTAATGATTTGTAAT 6180  
6121 CTTCTGATCATGAGCTTGTGCTACTGCTGCTTATTAATGCTCAATTAATGATTTGTAAT 6180  
6181 TTTACATGATATGATGATTTATTTTCCAGTTTATTAATATAGGCTATGAGCTATGTCAT 6240  
6181 TTTACATGATATGATGATTTATTTTCCAGTTTATTAATATAGGCTATGAGCTATGTCAT 6240  
6241 ATTGAAGACAGAAATCTGCTGCTTACCAAGTGGACAAATGTTTATTAATGATTAAGT 6300  
6241 ATTGAAGACAGAAATCTGCTGCTTACCAAGTGGACAAATGTTTATTAATGATTAAGT 6300  
6301 AATTGTAGAACAATAGGACTGCTAATCTCAAGTTCGCTGTGATGTCAAGTCAAGTCAATGTA 6360  
6301 AATTGTAGAACAATAGGACTGCTAATCTCAAGTTCGCTGTGATGTCAAGTCAAGTCAATGTA 6360  
6361 CAATTAACGTGATTTTCTCATATCTTTGATATCTGTTTACCTGTATGTCTTTTAA 6420  
6361 CAATTAACGTGATTTTCTCATATCTTTGATATCTGTTTACCTGTATGTCTTTTAA 6420  
6421 AGCATTTGGTGGAGTGTGATCCCTTTGATTTTAAATGATTAATGATTAATGATTTGGT 6480





misc\_feature 12156..12567  
/note="match to EST BG206034 (NID:g137272721)"  
misc\_feature 12157..12567  
/note="match to EST BG199786 (NID:g13721473)"  
misc\_feature 12366..12543  
/note="match to EST AW467230 (NID:g7037336) he07g06.x1"  
misc\_feature 12366..12499  
/note="similar to Homo sapiens EST N58568 (NID:g1202458)  
yvs5c09.r1"  
misc\_feature 12367..12539  
/note="match to EST BG204913 (NID:g13726600)"  
repeat\_region 13041..13097  
/rpl\_family="T-rich"  
repeat\_region 13075..13199  
/rpl\_family="L1"  
repeat\_region 13200..14202  
/rpl\_family="ERVK"  
repeat\_region 14203..14548  
/rpl\_family="L1"  
misc\_feature 14830..15092  
/note="match to EST AA005182 (NID:g1448797) zh93h1.r1"  
repeat\_region 15053..15274  
/rpl\_family="L2"  
misc\_feature 15413..15837  
/note="match to EST A1249019 (NID:g3844416) qh80h02.x1"  
repeat\_region 16526..16565  
/rpl\_family="AT-rich"  
repeat\_region 16932..17268  
/rpl\_family="MER1-type"  
misc\_feature 18420..19031  
/note="match to EST A1064746 (NID:g6359018)"  
repeat\_region 18715..18876  
/rpl\_family="MIR"  
repeat\_region 18934..18979  
/rpl\_family="MIR"  
repeat\_region 18980..19280  
/rpl\_family="Alu"  
repeat\_region 19281..19381  
/rpl\_family="MIR"  
misc\_feature 20122..20258  
/note="match to EST AW467230 (NID:g7037336) he07g06.x1"  
misc\_feature 20202..20258  
/note="match to EST AA425210 (NID:g2106002) zw47e06.r1"  
repeat\_region 20412..20761  
/rpl\_family="MALR"  
misc\_feature 21074..21452  
/note="match to EST AA860276 (NID:g2954271) ajs58e05.s1"  
misc\_feature 21234..21392  
/note="match to EST AA425210 (NID:g2106002) zw47e06.r1"  
misc\_feature 21234..21392  
/note="match to EST AW467230 (NID:g7037336) he07g06.x1"  
repeat\_region 21394..21436  
/rpl\_family="AT-rich"  
repeat\_region 21474..21511  
/rpl\_family="AT-rich"  
repeat\_region 21496..21681  
/rpl\_family="Alu"  
repeat\_region 21862..21968  
/rpl\_family="Alu"  
repeat\_region 21969..22101  
/rpl\_family="Alu"  
repeat\_region 24521..24613  
/rpl\_family="T-rich"  
repeat\_region 25161..25443  
/rpl\_family="Alu"  
repeat\_region 26345..26463  
/rpl\_family="L1"

Query Match 30.6%: Score 2011; DB 9; Length 160185;  
Best Local Similarity 99.9%: Pred. No. 0;  
Matches 2033; Conservative 0; Mismatches 0; Indels 2; Gaps 2;  
4536 GAAGATGAACAAGTTTCTGCTTTGAGGCACAGACTTTTCTGGAAGCAGAGCCAC 4595

|||||  
Db 55888 GAAGATGAACAAGTTTCTGCTTTGAGGCACAGACTTTTCTGGAAGCAGAGCCAC 55947  
Qy 4536 CTGAAGAGAGACACAGAAGAGCTCTGAGCATTTGAGACCTTGGAACTCACATTTCTGAG 4655  
|||||  
Db 55948 CTGAAGAGAGACACAGAAGAGAGCTCTGAGCATTTGAGACCTTGGAACTCACATTTCTGAG 56007  
Qy 4656 GACGCTGAGACAGTTTGGCTCTTCCCTGCTTTAAAGAGCATGGGGCTTTCTCCCC 4715  
|||||  
Db 56008 GACGGTGAACAGTTTGGCTCTTCCCTGCTTTAAAGAGCATGGGGCTTTCTCCCC 56067  
Qy 4716 TTCTTCTTTCCCTTTTGGCATGGAATCTGTGAAGAAATGCCCTGGCATTTCACA 4775  
Db 56068 TTCTTCTTTCCCTTTTGGCATGGAATCTGTGAAGAAATGCCCTGGCATTTCACA 56127  
Qy 4776 CTTTGTGCTTGAATGACACAGTGCAGCAATCTGAGCTCCACTGTGTCTGCTGCA 4835  
|||||  
Db 56128 CTTTGTGCTTGAATGACACAGTGCAGCAATCTGAGCTCCACTGTGTCTGCTGCA 56187  
Qy 4836 CATCACACAGTATCATTTCCAAATTCAGATCATCACAACAAGATGATTCACCTTGCTG 4895  
|||||  
Db 56188 CATCACACAGTATCATTTCCAAATTCAGATCATCACAACAAGATGATTCACCTTGCTG 56247  
Qy 4896 CACTTCTCAATGCTGGAAGATTTTAAATCTTCTTAAATTTAGATTTCAATCCAGTCT 4955  
|||||  
Db 56248 CACTTCTCAATGCTGGAAGATTTTAAATCTTCTTAAATTTAGATTTCAATCCAGTCT 56307  
Qy 4956 AGCAGCTGATCTCATTTGGGATATGAGAAAGCTAGACCATTAACACTTGGGGCTTTA 5015  
Db 56308 AGCAGCTGATCTCATTTGGGATATGAGAAAGCTAGACCATTAACACTTGGGGCTTTA 56367  
Qy 5016 ACCCACCAAGAGAGCAAGAAACAATGAATCCTTTGATGACAGTCTGCTGACCT 5075  
Db 56366 ACCCACCAAGAGAGCAAGAAACAATGAATCCTTTGATGACAGTCTGCTGACCT 56427  
Qy 5076 GTTTACAATGCTCCTCTTTT-AAAAAATAATGATTTAAAGATTTTTCAGAGATTA 5134  
|||||  
Db 56428 GTTTACAATGCTCCTCTTTAAAAAATAATGATTTAAAGATTTTTCAGAGATTA 56487  
Qy 5135 AATATATATCATTTTATGATTTATGATTTATTTAAACCTTAAGTAGGGTTCACGCT 5194  
Db 56488 AATATATATCATTTTATGATTTATGATTTATTTAAACCTTAAGTAGGGTTCACGCT 56547  
Qy 5195 GGTTCCTGAAAAACCAATATGCCGACAGGCTGTGCCACACCAAGAGAGGGAGAAC 5254  
|||||  
Db 56548 GGTTCCTGAAAAACCAATATGCCGACAGGCTGTGCCACACCAAGAGAGGGAGAAC 56607  
Qy 5255 CTGGCTGTGACCCGTGGCTTCCCATGCTCTTGGCTCACCGGCAAGTCCCTATCT 5314  
Db 56608 CTGGCTGTGACCCGTGGCTTCCCATGCTCTTGGCTCACCGGCAAGTCCCTATCT 56667  
Qy 5315 GGAAGTATGAAATGTTAGCAATTAATACCAAGACACCTCATCTGCTCCGCCAGTG 5374  
|||||  
Db 56668 GGAAGTATGAAATGTTAGCAATTAATACCAAGACACCTCATCTGCTCCGCCAGTG 56727  
Qy 5375 ATGGGCTTCTGTGAAACCTGTTTGCACATGCGCCAGGGAGGAACTGAGACCTTGTG 5434  
|||||  
Db 56728 ATGGGCTTCTGTGAAACCTGTTTGCACATGCGCCAGGGAGGAACTGAGACCTTGTG 56787  
Qy 5435 TCCTGTCTGAGCCTTATGAGGACGAGGAGTGTGATGCGATGCGTCTGCTCATTTG 5494  
Db 56788 TCCTGTCTGAGCCTTATGAGGACGAGGAGTGTGATGCGATGCGTCTGCTCATTTG 56847  
Qy 5495 AGATGATGCAAAACCCATTTTAATATATATTTCTTGAATTTGTTAATTTAGAGGT 5554  
|||||  
Db 56848 AGATGATGCAAAACCCATTTTAATATATATTTCTTGAATTTGTTAATTTAGAGGT 56907  
Qy 5555 GTAGGTTTGTGTTTG-TTTTTGTTTTTTTAAAGAAACATTTAATGATGATAC 5613  
Db 56908 GTAGGTTTGTGTTTGTGTTTTTGTGTTTTTAAAGAAACATTTAATGATGATAC 56967  
Qy 5614 ATTGAGTGAAGAGCTTGGGATGTGAGCTAAAGCCAGGTGTTATACGCTTTC 5673  
|||||

Db 56968 ATTGCAGTGAAGAACGACCTTGGGATGTTGGAGCTAATGCCAGCTGTTTACTGCTCTTC 57027  
Qy 5674 AAGACAGCCTCCCTTTATTGAAATTGGCAATTAGGGAATAAACAAGCCTTTAAACGTGATA 5733  
Db 57028 AAGACAGCCTCCCTTTATTGAAATTGGCAATTAGGGAATAAACAAGCCTTTAAACGTGATA 57087  
Qy 5734 AAGATCAAAAACCTGGTTAGACATGGCAGCCTTTGCAAGGCGTGAAGTACCAAGACAT 5793  
Db 57088 AAGATCAAAAACCTGGTTAGACATGGCAGCCTTTGCAAGGCGTGAAGTACCAAGACAT 57147  
Qy 5794 AACCTCAAGTGGCTTTATGAGAGCTGCATATAGAGAGGCTTAAGTGAAGCAACATCT 5853  
Db 57148 AACCTCAAGTGGCTTTATGAGAGCTGCATATAGAGAGGCTTAAGTGAAGCAACATCT 57207  
Qy 5854 GCTCAGAGCTGCTATTATACCTTATATGACTGAATAGCCCCCTCATCTATTGTTGTGT 5913  
Db 57208 GCTCAGAGCTGCTATTATACCTTATATGACTGAATAGCCCCCTCATCTATTGTTGTGT 57267  
Qy 5914 TGTTTTGCACAGAGCTCGGAAAGTGAAGGCTGCCAATCTGAGTACTCAAAATGTGAG 5973  
Db 57268 TGTTTTGCACAGAGCTCGGAAAGTGAAGGCTGCCAATCTGAGTACTCAAAATGTGAG 57327  
Qy 5974 GAACCTGCTGCTGGATTTTTCATTAAATTCAGTGCATATGATGATGATGATGAT 6033  
Db 57328 GAACCTGCTGCTGGATTTTTCATTAAATTCAGTGCATATGATGATGATGATGAT 57387  
Qy 6034 AAACGTAATAGCTCAAAATTTTAAAGTGAATTCGACGCTTTTTCACCTGATCAAC 6093  
Db 57388 AAACGTAATAGCTCAAAATTTTAAAGTGAATTCGACGCTTTTTCACCTGATCAAC 57447  
Qy 6094 AATGTAGTCTTATTTTAAATTTCTCTGATCATGAGCTTTGTCATCTGCTTAT 6153  
Db 57448 AATGTAGTCTTATTTTAAATTTCTCTGATCATGAGCTTTGTCATCTGCTTAT 57507  
Qy 6154 TACATGTCATTAATGCAATTTGTAATTTTACATGTAATGCAATTAATGTCAGTTTAT 6213  
Db 57508 TACATGTCATTAATGCAATTTGTAATTTTACATGTAATGCAATTAATGTCAGTTTAT 57567  
Qy 6214 TATATAGCTATGAGCTCATGTCATATGAAAGACAGAAATCTGCTTACCAAGT 6273  
Db 57568 TATATAGCTATGAGCTCATGTCATATGAAAGACAGAAATCTGCTTACCAAGT 57627  
Qy 6274 TGCACAATGTTATCTTAAGCAATTAATGTAAGCAATAGGACTGCTAATCTCAAGTTC 6333  
Db 57628 TGCACAATGTTATCTTAAGCAATTAATGTAAGCAATAGGACTGCTAATCTCAAGTTC 57687  
Qy 6334 GCTCTGTGATGTCAGAGTGAATGTAATTAATGCTGATTTCTCATACTTTTGATA 6393  
Db 57688 GCTCTGTGATGTCAGAGTGAATGTAATTAATGCTGATTTCTCATACTTTTGATA 57747  
Qy 6394 CTACTTGTACTGATGTCATTTTGAAGAACATTTGGTGAAGTCTGATTTGCTTAT 6453  
Db 57748 CTACTTGTACTGATGTCATTTTGAAGAACATTTGGTGAAGTCTGATTTGCTTAT 57807  
Qy 6454 TTTAATCAATTAATGTAATTTGTTATTTTGTGAAGTGTGTAAGTATGATA 6513  
Db 57808 TTTAATCAATTAATGTAATTTGTTATTTTGTGAAGTGTGTAAGTATGATA 57867  
Qy 6514 TGTATTAGCTTCTACATCCAGTTTGTACAGCTGGGAAATTAATTAATTAACAT 6568  
Db 57868 TGTATTAGCTTCTACATCCAGTTTGTACAGCTGGGAAATTAATTAATTAACAT 57922

RESULT 5  
AF394782 5742 bp mRNA linear PRI 05-NOV-2001  
LOCUS AF394782  
DEFINITION Homo sapiens rap guanine nucleotide exchange factor mRNA, complete cds.  
ACCESSION AF394782  
VERSION AF394782.1 GI:15077825  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
21551169  
11524421  
2 (bases 1 to 5742)  
Geo, X., Satoh, T., Liao, Y., Song, C., Hu, C.D., Kariya, K. and Katoaka, T.  
Identification and characterization of RA-GEF-2, a Rap guanine nucleotide exchange factor that serves as a downstream target of M-Ras  
J. Biol. Chem. 276 (45), 42219-42225 (2001)

CDS

183..4712  
/note="RA-GEF-2"  
/product="rap guanine nucleotide exchange factor"  
/protein\_id="AK83368.1"  
/db\_xref="GI:15077826"  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
translation="MNSPYDGRARALAKKRPERRPDLNTYSILHGMETLSNIREH  
QLRMSARARYERSGNQVLFCESTIAKCMVILSGVLYVSGMVLPCPSGKRGK  
RGCDLVEPESEMIIVENAKDNESILLOREIPARSRFRKINKGRQRTIDVEY  
NSYLSPADLTLMHLTENPHOVYVSSSGGCSIASDSGSSLSLDIOAESEVGDV  
DLTRLEPGVSEDEDEDEEDIDRDPLGGRLVCELEKPEADTDODIQLLEFHM  
OLPARANMVSARELCVMIJFEVVEQAGAILIREGDELDSMYVLINDGILSHDCK  
VENLPMGNSFGTPTPLDKOYMHGIVRTYVDDCPICINQOQYMRILNVEKNTKVE  
BGEIVAVHEHRLDLSGRKHIVIKAPERLIMHLEHSIVDPTTIEDLRTYRTE  
LESPLDVGIKLEMPKIDSLRDQVTRIVLWVNNHFNEDDPATRTLEEFNKLED  
TKMNGHLRLNLJACAKAKRWVYLOKASRESLPDPSLNGSEKGFGLVEGVEKQ  
AADSGLKRDQDINEYVGNFENITFMKAVELIRNTHALTYKTIPEFKLLEFTEQ  
EKSIVPHIKTAEKKSNNRSIOHVGDIQSOEKSQKSVKANTVSGGRNLIIRIKLKD  
TRESILPKLEFSDGSLSSQSDSYGSTRCHRSILAIMPILPTGLSSSPDLIOPTSM  
DPSNPSAVGEYITPQVIRKRVDOOSCTIIISKTTAKVEYFVAVHEFGLTGASDY  
SLCESVYPEGVIKQRLPDEFSKLADRIQLNGRYLLNNMETETLSCDEDAQELVKE  
SOLSMOJLSTIEVATOLSMRDEFLRNIEPEYIDDLKLNKSGTNTLHKKREIDVNO  
ETFWVASLITFEANOLKRMKLIKHEIKTALHRECKNNSEFALISGLNLASVRLRG  
TWEXLPSKYEKHLQDLPDPSRNWAXYRNILSSQMPPIIPLFVKKADMPFLHE  
GNDSDVGLVNEFKLRMTSKETROYVVRNANMDPAMFMRKKRMRSLGSLSSQSTN  
SNMLDVGSAHKKRRARRBSLNAKKLYEDAOAKRKYOLSLDYEMPEERFOMSLIO  
WEPAYGTLTKNLSEKRSKSSSEMSVPMKSAQITTKAHLDHQRHSYLOYPAYVHLH  
IRKKQTKDPAINTSLPQKVLGTTEIEISGKKTEDTIVASLSHSPASVQSGFHG  
YVLISAKSDNLSDSHSEISRSISVSNCSVDSKSNALODERSSOALAVDSGAL  
EKTEHSAIGDISHOGPCWTLKPLKICLAWSSVSENEISOEHIILEADSGSGSM  
TSCSSSHDNFQSLPNPKMSDFLNSRYRTHDDPFAVEEPTDSBEYSCKSKSCRFQCG  
CKGSLERKMSWSSSLSDTIEPNYGTVARVLESLIPASSSELDKNDTDVYKNTVS  
STKGLIENQVSAV"  
BASE COUNT 1757 a 1091 c 1323 g 1571 t  
ORIGIN  
Query Match 24.5%; Score 1612; DB 9; Length 5742;  
Best Local Similarity 66.2%; Pred. No. 0;  
Matches 2548; Conservative 0; Mismatches 1210; Indels 90; Gaps 12;



[illegible]

Df	3966	GTTACACACTTATTCCTCATGAGTAATCTGTCAACAATTGCCTCCAGCATACTGGAG	4025
OY	3504	ATTTCCTCAGCATTCAGTATTGTTTAGCAATTCGTCCTTTTGACTCAGTCCAGTCTCACTG	3563
Df	4026	ATTCTTCACCGCTCAGCATCGTGAAGAATGTCTGTTGACTCCAAATGCTCGACGCTCTA	4085
OY	3564	CACCATGAGAGGCCCAAGGCAATTCCTCTCAGCATTCGTCGGAACAACAACCTAGSATTGGCC	3623
Df	4086	CAGGATGAAACGGTGTTCCTCTCAGGCCCTGGCACGTCCCCTGAATCCACTGGGGCAATTGGAA	4145
OY	3624	AGGATGAGAGAGCGGACCATGATTGCAACCTGATCATGTTACTCTGGGGTCCCTATGACCA	3683
Df	4146	AAGAACAGAGACACCCCTTCAGGGGATAGAGATATATAGTCAACATGGCCCTGGSTGGACATC	4205
OY	3684	ATGTCGGAAGGCCAGGCGCTTATATGCTACAGCTACAGTAAATTCCTTCACAGCAGAG	3743
Df	4206	TTAGAGCCATCTCTATCAAGATGTTTACTGTCTCA-----TCGTCTGTAGCAAATGAA	4259
OY	3744	GAACTTTCCCAAGATCAGGGGATTCGGCCGTCACCTGATGCTGTGCAGAGGCGCGTGGG	3803
Df	4260	GAGATTTTCCAAGACCAATATCATTT-----ATAGAAGAGCCTGACAGTGTCGTGGA	4310
OY	3804	ACCTGAGAGCTCATGCTCAAGTGGCTCCCATGATATATATACAGAGATCCAGACACAGGA	3863
Df	4311	ACTTGGACTCTGTCTTCAAGCAGCTCCCATGATGACCACTTCCAAAGCCTTCCAAACCCAAA	4370
OY	3864	AGCTGGGA 3871         	
Df	4371	AGCTGGGA 4378	
RESULT 6	AF478567	6390 bp mRNA linear PRI 02-MAY-2002	
LOCUS	AF478567	Homo sapiens PDZ domain-containing guanine nucleotide exchange factor PDZ-GEF2 mRNA, complete cds.	
DEFINITION	AF478567.1 GI:20386205		
ACCESSION	AF478567		
VERSION	AF478567.1		
KEYWORDS	.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 6390)		
AUTHORS	Banville,D., Murthy,K., Shen,S., Clark,K. and Fortin,Y.		
TITLE	A PDZ domain containing guanine exchange factor (GEF) interacts with the second PDZ domain of human PTPLE		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 6390)		
AUTHORS	Kozlov,G., Banville,D., Gehring,K. and Ekkiel,I.		
TITLE	Solution structure of the PDZ2 domain from cytosolic human phosphatase hPTPLe complexed with a peptide reveals contribution of the beta2-beta3 loop to PDZ domain-ligand interactions		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 6390)		
AUTHORS	Banville,D.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-JAN-2002) Biotechnology Research Institute, 6100 Royalmount, Montreal, Quebec H4P 2R2, Canada		
FEATURES	Location/Qualifiers		
source	1..6390		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="5"		
	/map="5q31.1"		
	207..5012		
	/codon_start=1		
	/product="PDZ domain-containing guanine nucleotide exchange factor PDZ-GEF2"		
	/protein_id="AM21637.1"		
	/db_xref="gi:20386206"		
	/translation="WNSPVDDGARQALRKRPETPEDLNTIYSYLHGMLISULREHDLKRNIAARERYSGNQVLCSEFIACWVILLISGVLVKGSWVDPCSFQFGGRK"		

RGCDLVLEPSEMIVENAKNDNEDSILOREIPAROSRRFRKINXKGERQITTDVEV  
NSYLSIPADLFKMLHTENPHQVTHVSSOGCSIASDSGSSLSIDIQATSESEVGD  
DUTRLPEGEVDSDEDEDEEDPDTDFPLOGDIAVECELEKPEAKDDDDIOLLEFMH  
OLPAPANTMSVRELICVMIPEVEQAGAIILBEGOLDSMYILNKTVEISHPDCK  
VENLEPMSFEITPLDQOMHGIKRVIVDQCFPCIAQODWRLINVENTHVEE  
EGEIVMHEHEHRELDGSRKHIVIKATPERIMLIEHSIVDTYIEEDLYRT  
LESPLDVGILKLEWFKIDSLDKVRIYLLWNHNFEDPANTRLLEFEKLEED  
TKMNGHLKRLINACAKAKROVYLOKASRNSWPNFANGSEKGFJFVGEVSGK  
AASGLKRGDDIMEYNGONFENITFMKAVELIRNNTKIALTVKTLIPFKELPTED  
EKSGYPIHPKIAEKSNHSHONHPGDILOEIOSEKSKVKVANTYSGRRKTRILDK  
TRFSLIPPKLFSDGLOSODSIVGTHCHRSIATIMP1PGTSSSPDLQPTTSM  
DEFNSPDIPOVIRVFKYDOQSVIISKDTYAKVYHVAHEPGLTCAELYSLEY  
SVTEGVIKORRLPOEFSKLADRIOLNREYTLKMNEMETIESDPAQADYSOLSM  
LOLSTIEVATOLSMRDFLFRNIEPTEYIDDLFNKSTNGKILREFDIQOETFMV  
ASELITEANOLKRMKTIKFKIALHCECEKNEFMFIISGLINASARILNGETEKL  
PSKTEKHLODIDITDPSRNMKTRNIISSOSMOPPIIPLPVAKMDTFLEHGDNSK  
VDGLVNEFKRLMISKEIROVAVMTSANDPAMPEPMEFOSGTSINMLDVOGAKHK  
RARRSLINAKKLYEDAOARKVYOYLSLVEDEEKFOMASLOMEPAVGTITANLS  
EKRSKSSSEMPVPRMSAGQTTKAHLHOPHRISOVLOVPAVNLHPIRKQGTDPALN  
TSLPKVLTTEIEISGKHTEBDTISVASSLHSSPASOGSPHKGYTLIPSAKSNLS  
DSHSSEISRSISVNCSDVDSASAILOPERSOALANPESTGALEREHAGIGDHS  
OHGPMVTLKPSLIKCLAVSSSVNEEISQEHITIEADSGSGWTSKSSSHDNFOS  
LNPKSWDFLNSYRTHLDPIAEVEPDPSEBYSCKSKSRFCQCKGLERKSWTS  
SLSLDTYEPNGTVKRVLESTPAESSEGLDPKADDPVYKTVISSTEGGLIVCVTS  
PKKDRYREPPPTPGYLGISLADLKEGPHLKPDPDYSVAORSKMMHNSLSRLPAP  
SLSSNLVACVPSKIVTPORHNLQPFHFKLGVDVADSEADENEGVANS"

BASE COUNT 1950 a 1242 c 1447 g 1751 t

ORIGIN

Query Match 24.1% Score 1585.4; DB 9; Length 6390;

Best Local Similarity 66.0%; Pred. No. 0;

Matches 2528; Conservative 0; Mismatches 1206; Indels 99; Gaps 12;

QY 45 ACCATATGTTTCTTCTATATGAACACATAGCATCCAGCTRACCATGAGTTATGGGC 104  
DB 624 AGAAGATTTTCGAAATTTACTATAAGAGAGCCGCAACATTTACTATGATGAGG 683  
QY 105 CAGCAGAGAAACACTCATCTCTCGACATTTACAAAACCTGATTTCTGACAGTCTC 164  
DB 684 GTTAAACAGTATCTTCTCTCCAGCTGATCTTACCAATGATGATTCACAGAAACCTT 743  
QY 165 CACCCACAGTGACCCACGCTTTCTTGTAGCCATTCAGATGATGATTCATGATGATTC 224  
DB 744 CATCCACAGTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 803  
QY 225 GGGAGCAGCAGTCTTCTATATCTACAGGCCACAGAAAGCAGGCTGTGATGAC 284  
DB 804 GGAAGCAGCAGTCTATCTATCTATCTACAGGCTACGAGAGTGAGGATGATGAT 863  
QY 285 CTGAGTGGTGGCAAAACAGCAGATTCGAGAGAGAGCAGCAGCAGTAAAGACATTT 344  
DB 864 TTGACACGCTTCCAGAAAGACCTGTTGATTCGAGAGATGAGAAAGAGAGAGAG 923  
QY 345 GAGAGAGCATCAGATCTCTGATGAGCAGGACATTTGTAGAGACTGCTTAGAGAGAC 404  
DB 924 ATTGATCGAAGATCCATTTGAGGGGCGAGATCTTTGCGAGAATGCTTGAAAAAGA 983  
QY 405 CCAATTGACCGGACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 464  
DB 984 CCGTAGCAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1043  
QY 465 GCTTTTGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 524  
DB 1044 GCTTTTGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1103  
QY 525 GTGTGTGAAGAGCAGGACCATAGTGAATGATGATGATGATGATGATGATGATGATG 584  
DB 1104 GTGTGTGAGCAGGCTGAGCTATATTTGGAAGATGGGCAAGAGCTTGAGCTGATG 1163  
QY 585 GTGATTTCTCAATGATCTGTGAAGTGAATTCAGATGAGAAAGCAGAAATACTGTGC 644  
DB 1164 GTTATTTTAAAGCGCAGCTGTGGAATCAGTCCAGATGAGAAAGTGAATTTGTTT 1223

QY 645 ATGGCAATAGTTTGGTGTCTCTCTACCAAGACCAAGAAATACATGAAGAGATGATG 704  
DB 1224 ATGGCAATAGTTTGGTGTCTCTCTACCAAGACCAAGAAATACATGAAGAGATGATG 1283  
QY 705 AGACAAAGGTGAGTCTGCAATGATGATGATGATGATGATGATGATGATGATGATG 764  
DB 1284 AGGACTAAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1343  
QY 765 CTCATCAAGTGAAGAAACATGCAAAAGTTGAAGAGAGAGAGAGATGATGATGATG 824  
DB 1344 TTAAACCATGTGGAAGAAATACCCCAATGAAGAGAGAGAGAGATGATGATGATG 1403  
QY 825 AAAGAACCCGGAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 884  
DB 1404 CATGACATCGGGAATCAACCGGAGTGAACCCGAGAGAGAGAGATGATGATGATGATG 1463  
QY 885 ACCTCGAAGAGTTAAACATGCAATTTGTTGGAAGAGACATTCAGTATGATGATGATG 944  
DB 1464 ACACCTGAGCGTCTCATTAATGCAATTAATGAAGAAACATTCATGCTGATGATGATG 1523  
QY 945 ATGAAAGACTTCTGTGACCTATAGACTTTTCTTTAGCCCAATGCAAGTGGCCAAA 1004  
DB 1524 ATGAAGATTTTCTATTAACCTTACAGACATTTCTGAAAGTCTTGTGATGATGATG 1583  
QY 1005 AAGTTATGAGAGTGTATTAATGACCCGAGCTCAGGATTAAGTTACAGCGATGATTA 1064  
DB 1584 AAATCTTTGGAATGTTTAAAGATCGACAGCTTAAAGATTAAGTTGACACGATTTGATTA 1643  
QY 1065 TTGTGGTAAATTAATCACTTCAATGATGATGATGATGATGATGATGATGATGATGAT 1124  
DB 1644 TTATGCTAAATTAATCACTTCAATGATGATGATGATGATGATGATGATGATGATGAT 1703  
QY 1125 GAAAGATTTGAACATCTGGAAGAGAAATGGGTGAGACACTTAAGCTGTGAT 1184  
DB 1704 GAGAGATTTGAACATCTGGAAGATGACAAAGATTAAGTATGATGATGATGATGATG 1763  
QY 1185 ATCCGCTGTGCTGCTAAACCAAAAGAGATGATGATGATGATGATGATGATGATGATG 1244  
DB 1764 ATTGCTGTGCTGCTAAACCAAGCTTAAGTGAAGAGATGATGATGATGATGATGATGATG 1823  
QY 1245 GCTCTTTGCTTTTATCTTACTTGAAGCTGTGAGAGAGATTTGATGATGATGATGATG 1304  
DB 1824 TCCCTCTCAATTAATGACCTTAATGAGAGAGATGATGATGATGATGATGATGATGATG 1883  
QY 1305 AGTGTGATTTCAAGTACCAAGCAACTGAAGCAGCTTGAACGGGGGATCAGATTTA 1364  
DB 1884 GAGTGAACCTGTGCAAAAGCTGCTGATTCAGACTGAAAGCTGTGATGATGATGATG 1943  
QY 1365 GAAGTAAATGGCCAAACCTTGAACATTCAGCTGCAAAAGCTTAAGAAATTTCTTGA 1424  
DB 1944 GAAGTAAATGGCCAAACCTTGAACATTCAGCTGCAAAAGCTTGAAGAAATTTCTTGA 2003  
QY 1425 AATTAACACATTTATCTATCACTGTAACCAATTTATTTATTTAAGAACTTTA 1484  
DB 2004 AATTAATCACTATCTGACTTACTGTAAGACCAACATTTTGTGTTCAAGAGTACTT 2063  
QY 1485 ACAAGATTTGACAGAGAAAGAAATGGTCCGCCACCTCTTAATTTGTGATGAT 1544  
DB 2064 TTATG---GACTGAAACAAAGAAATCTGGTGTCTCATATTTCCAAAATTTGCGAAAA 2120  
QY 1545 AAAAAGGCCGACCTGATCCAGATCTGCGTGTGATGATGATGATGATGATGATGATGATG 1604  
DB 2121 AAAAGT---AATGCAATTTCTATCCAGCATGTGCCAGAGATATTTGAAC---GACATCA 2174  
QY 1605 CTTGAAAGAGTGAACAAAAGTAAAGCCACACCTGTG---GGAGAGAGGAACAGCTG 1661  
DB 2175 CAGAGAAAGGAAGTGAAGAAAGTTAAAGCAATACGTTCAGTGAAGAGAAACAAATC 2234  
QY 1662 AAAAATATCTGCAAGACTCGGATCAGTATCTTCCACAGAAACATTCATGATGAT 1721  
DB 2235 AGGAAGATTTTGAATAAACACGATTTAGTATCTTCCACAAAGCTATTTAGTATGGA 2294  
QY 1722 GGGATTTGGTCACTCAAGATGACAGCATGATGATGATTAAGGACGAAAGCAATCCCA 1781

Db 2295 GGCCTAAGCAATCACAAAGATGACACATTTGTGGAAACAGCAGTGTAGCGCTACTGTC 2234  
QY 1782 ACTGCATTGCGCTGTCAGTGGAACTTATCATCCAGTAATCTGATTTTANTTGCAGTCACAT 1841  
Db 2355 GCTATATAGCCCTCCCTGACACACTCTTCATCCAGCAGCCCTGATCTCCGACGCTTACC 2414  
QY 1842 CATCGATTTTACATTCAGTGTCTACTCCGATTCGCGAGATTCGAGTGCATAGCGGTTT 1901  
Db 2415 ACCAGTATGATGATTTTTCATTCCTTCAGATATCCCTGATATCCCTGATTAAGATTTTC 2474  
QY 1902 AAGCGATGACGAAAGCCCTACATCATGATCAGTAAGGACACTACGAAAGAGATG 1961  
Db 2475 AAGTGATGACGAAAGTTGCTACATATCATCAGTAAGACACACAGTAAGAGTA 2534  
QY 1962 GTCATTGAGCTATCAGAGGATTTGCTGTTACTGCGACCCCGATCATATTTACTATGT 2021  
Db 2535 GTTTTTCATGCTGTTCATGATGTTGGTTGGACCGGTGCATCCGACATATTCCTCTGT 2594  
QY 2022 GAGTCTGCTGACCTGAGGAGTAATCAACAAAGACATTCAGATCAGCTTTC 2081  
Db 2595 GAAGTTTCTGTACTCCTGAGGTTGTCATTAACAGAGAACTTCAGATCAGTTCTCC 2854  
QY 2082 AAACCTTGACAGACAGAAATACAACTAGTGAAGTATATCTGAAAAACAACATGAAACA 2141  
Db 2655 AATATGACGTATAGATTCACATCACTCAATGGAAGTATTTACTTAATAATACATGGAACA 2714  
QY 2142 GAACCTCTTTGTTACAGATGAAGATGCTCAGAGATTGTTGAGAGAGTCAATTTCCCTC 2201  
Db 2715 GAACCTTATGTTTACATGATGAAGATGCTCAAGAACTAGTTAAGGAAAGCCAGCATCCATG 2774  
QY 2202 CTTCAGCTGACGACTGGAGGTTTGCACACAGCTCTCTTGGAAATTTTGAACCTTT 2261  
Db 2775 CTCGAGCTGCTACCTAGTTGAGGTGGCCACCAGCTGTCAATGAGGACTTTGATTTGTTT 2834  
QY 2262 CGCAACCTTGAACCTACTGATATATAGATGATTTATTTAACTCAGATCAAAAACAGC 2321  
Db 2835 CGTAAATATGAACGACTGAGTACATGATGATGATTTTAAAGTTAAATTCCAAACAGCA 2894  
QY 2322 TGTGCCACTGAAAGATTTGAAAGAGTCAATTAACGAGAAACATTTTGGTAGCTCT 2381  
Db 2895 AATATCTATTTTGAAGGTTTGAAGACATTTGTAACCAAGAACATTTGGTTGGCTCA 2954  
QY 2382 GAATTTCTCAGAAACAACACAGCTGAAGAGATGAGATTCATTAAGCATTTTCATCAAG 2441  
Db 2955 GAATTTTACTGAGCAAAATCAAGCTCAACGATGAAGATTTATAGCATTTTATTAAG 3014  
QY 2442 ATAGCAGCTGAGTGAAGATGCAAGATTTTACTCAATGTTTGCATCATCAGTGGC 2501  
Db 3015 ATTGCATTTCTGTCAGATGTAAGAACTCAATTCATGTTTGCATTAATTAAGTGGC 3074  
QY 2502 CTAAACCTGGACACAGTGGCAAGACTGCAACACCTGGGAAACCTTCCCATTAATAC 2561  
Db 3075 TTGAACTGGCATCTGTAGCAAGACTGAGAGAACTTGGGAAAGTTTACCAAGCAATATC 3134  
QY 2562 GAAAACTATTTCAAGATCTCAAGACCTGTTGATCTTCCAGAAACATGCGAAAAATAT 2621  
Db 3135 GAGAAACATCTTCAAGATCTCAAGACATTTTGTATCTCATCTGAACATGCGAAAAATAT 3194  
QY 2622 CGTAATGTTCTCAATTAATCTCAACCTGCTCAATTCCTCTATTTCCAGTTATC 2681  
Db 3195 AGAAATATTTCTGTAGTACTCAAGATGACAGCTCAATTTATTCACACTTCCGTTGTC 3254  
QY 2682 AAAAGGATGACCTCTCCCTTCAGAGAAATGATCAAAAGATGAGGCGGTGCTCAT 2741  
Db 3255 AAGAAAGATATGCAATTTCTACATGAAGAAATGACTCCAAAGTAGTGGTTTAGTAAC 3314  
QY 2742 TTGAGAAGCTAAGATGATTGCAAAAGAAATTCGTACGCTTGCGCAATGCTTTCAGT 2801  
Db 3315 TTGAGAAGATTAAAGATTTTCCAAAGAAANTCCGCAAGTGTTCGAATGACTCTGCT 3374  
QY 2802 AACATGACCTTGCCCTCATATTTCTAGAGACTGGAAGAAAGATTTGGGCTCT 2861

Db 3375 AACATGACCCAGCATATGATGTT-----TCGACAGAGTCA 3410  
QY 2862 CTCAGCCAGGGTAGTATCAAAATGACAGTGTAGATGTTGCTCAGACAGTGTGATATAA 2921  
Db 3411 CTGAGTCAAGGAAAGCAAAATTCAAACATGCTGAGAGT---TCAGGAGAGTGTCTCAAAA 3467  
QY 2922 AAGCGGTACGTCGTATGTTCTCTTCTCAATGCCAAAGATCTTTTAAAGATCCCAATG 2981  
Db 3468 AAAGGGCAGCCCGCAGCTCTCTGCTTAATGCCAAACACTATATGAGATGCCCAATG 3527  
QY 2982 GCTGAAAAGTGAAGCAGTACCTTTCCAAATTTGAGCTAGAAATGAGCAGAGAGTCTT 3041  
Db 3528 GCAAGAGAGTGAAGCAGATCTTTCCAGTCTCAGTGTGAGCAGAGATGAGAGAGTTC 3587  
QY 3042 CAGACATTTATCTGTGAGTGTGAGCCAGAACCAACACATTTGCTTAAGATCCGTGTGAC 3101  
Db 3588 CAGATATGTCATTTACAGAGGAGGCTGGAATGTATGTTACTTGACCAAGATTTAAGTGA 3647  
QY 3102 AAAAGCCTGTCA---ATCCGAGACCTCTCCAGTACGTCCTCAAGGCGAGGTGCACAAAG 3158  
Db 3648 AAAAGATCAGCCCAATCATCTGAAATGTCTCCAGTGTCTATGAGTCTGCGCCAAACA 3707  
QY 3159 AAGCTGATCCCTGCGACAGCCCGCAGCAGCAGCAGCAGCAGCATATAATCAACAG 3218  
Db 3708 ACTAAGCCCATTTGATCAACCC-----CACAGATGAAGCCAG 3746  
QY 3219 GAGCTACAGTTCGCCCGCTGTCCTTTATCTTACAGGAAGAAAGTGGCCGTAAGAGAT 3278  
Db 3747 GTCCTTCAGGTGCGACTGTAAATTTGACACCCCATCAGAGAAAGGAGAACAAAGAGC 3806  
QY 3279 CTCGCCATTTTGGCATTAACCTCCACAAAGCTTTAAAAAAATTTCTTTTGTGTGAA 3338  
Db 3807 CTCGACATGATACAAAGTTTACTCAGAAAGTTTGAAGAAC---CTGAA 3854  
QY 3339 GAAGAAATTTGGAAGCTCAAGAAACAGGCTGAAGATATCAATATCAATGATCTTGG 3398  
Db 3855 GAATATA-----GTGTGAAGAGCATACGAAAGACATATTTCTGTGGCTCATCT 3905  
QY 3399 CAGCTTCT 3458  
Db 3906 TTATCTTCTAGTCT 3965  
QY 3459 CCCAGTGTACTGTGATAATTTTTCAGATTTCTGTCAGAGTGAATTTCTTCAGATCC 3518  
Db 3966 CCATCAGCTTAATCTGACAACTGCTCTGACTCCAGCATATGAGATTTCTTCACGCTCC 4025  
QY 3519 AGTATGTTAGCAATTCGTTTGTGACTCAGTGCAGTCTCACTGACATGAGAGGCC 3578  
Db 4026 AGCATGCTGAGCAATTTGTTGTTGACTCCATGCTGCAAGCTCTTACAGATGAACGCT 4085  
QY 3579 CAGAGCATTTCTGTCAAGCATCGTGAACAAACCTAGGGATGGGCGAGATGAGAGCGG 3638  
Db 4086 TCTCTTCAGGCCCTGGCAGTCCCTGATTCACGTGGGSCATTTGAAAGACAGACGCT 4145  
QY 3639 ACCATATTTGAACCTGATAGTATGCTTGGGCTCTATGCTACCAATGTCCAGGCGCA 3698  
Db 4146 TCAAGGATAGAGATATATATCAACATGCGCTGGGTGACACTTCTTGAAGCATCTTA 4205  
QY 3699 GCGTTATGCTTACAGCTATAGTAATTTCTTCCAAAGCAGAGAAAGACTTCCAGAGAT 3758  
Db 4206 ATCAAGTGTAGCTGTCTCA-----TCGTCGTGAGCAATGAAGATATTTCTCAAGAG 4259  
QY 3759 CAGGGGATGCGCGTACACTTGATGCTGTGACAGTGGCGGTGGAGCTGAGCTATGCG 3818  
Db 4260 CATATCATTT-----ATAGAAAGCAGCTGACAGTGTGCTGTAAGTTGCTGTGT 4310  
QY 3819 TCAAGTGTCTCCCATGATATATATCAAGAGCATCCAGCACGAGAGCTGGCA 3871  
Db 4311 TCAAGAGAGTCCCATGACACTTCCAAAGGCTTCCAAAGCCCAAAAGCTGGGA 4363

RESULT 7  
AF117947

LOCUS AF117947 6401 bp mRNA linear PRI 01-JAN-2000  
DEFINITION Homo sapiens PDZ domain-containing guanine nucleotide exchange factor 1 mRNA, complete cds.  
ACCESSION AF117947  
VERSION AF117947.1 GI:6650765  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE 1 (bases 1 to 6401)  
AUTHORS Kawasaki, H., Chen, E.J., Springett, G.M., Graybiel, A.M. and Housman, D.E.  
TITLE A new family of Rap guanine nucleotide exchange factors  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 6401)  
Housman, D.E.  
Kawasaki, H., Chen, E.J., Springett, G.M., Graybiel, A.M. and  
TITLE Direct Submission  
JOURNAL Submitted (06-JAN-1999) Center for Cancer Research/Dept. Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA  
FEATURES  
SOURCE Location/Qualifiers  
1. 6401  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
70. 3684  
/note="PDZ-GEF1"  
/codon\_start=1  
/product="PDZ domain-containing guanine nucleotide exchange factor 1"  
/protein\_id="AAAF22004.1"  
/db\_xref="GI:6650766"  
/translation="MAPTLFQKFSKRTGIGAPDARDPDCGSEMPLEDPDSQIRL IYQDCERGRNVLEDSYKRRNEDISVDLNTITSYLHGMEITLSNLEHDLRLSAR ANTERYSQNVLFCESETIARQYIILSGSVLVPCSGQFGKRGKDCDLVA EPSEMIYVENAKDNEDSILQREIPARQSRRRKINYGKREITIDDEVNSYSLA DLTKHLNLEPHQYTHVSSQSGSISADSSPSDIQATSESEVDVLTIRPEG PVSDEDEDEIDRTDPLQGRDLVRCLKEPEKDDDDIOLLEFMHOLAPAFAN TMSVRELCSWMEFEVBOAGAILLEDOQELDSVYIINGYEISHPGKVENLEMGN SFGITPTLDKQIMGIVRTKYDDCOFVCIAQODIWRILNHEKNHKKVEESELVHW EHRELDKSGTRKHIVIAATPERLIMHLEHSIVDPYIEDFLITRTFLESPLDVG ILMLEFMTIDSLRDVTRIVILVMNHNHFNDEGDAMTRFLEFEKNLEDTYMGHLR LILNACAKAKARQVYLOKASRESPLQSLNGSKGGEIYEGEKNLEKSKADSLKR GDOIWEVNGONFENTTFMKAVELILNNHILALTVTNIFVEKELLFRTEQSGVPLP PIAKKSNRHSIOHVPDIDIEPTSEKSKKATVYGGRRKIKILDKTRFSLLP KLFSGGLSQSODDSIVTGRHRHSLAIMPIDGLTSSSPDLQPTTSMLEDSNSDI PDQVTRVERKVDQSCYIIISKDTAKKEVFAHVEGLGADVYSLEVSATPBGVI KQRLPDPFSKLADRIQLNGRYLLKNMETETLCDEDEQAEIKVESQLSMQLSTIEV ATQLSMRPDELRIALTEPEYIDDLFKLSKNTGNTILKFEEDIVNGETFWASIELTEA NOLKRMKLIKHFIKIALHCBECKNFNSMFAIISGILASVARIETWELPKSKYKHL NOLDIIDPDSRMAKYRNITLSOSMOPILIPFPVKKDMFPHGNGNSKQGLVNF EKLRMISKEIRQVRYRNTSANMDPAMFPROBSLSQGSTNSMMLDVCGAKHKKRRBSLL NAKKLIYDAQARKVKOYLLSLDVLDEDEKFOUMSLQWEPAYGLTKNLSEKRSKII "

BASE COUNT 1971 a 1236 c 1435 g 1759 t  
ORIGIN

Query Match 24.1% Score 1584.4 DB 9 Length 6401;  
Best Local Similarity 65.9% Pred. No. 0;

Matches 2528; Conservative 0; Mismatches 1206; Indels 100; Gaps 12;

QY 45 ACCATATGTTTCTTCAATATGAACACATAGCAATCCAGTAAACCATGAGTATGAGGC 104  
DB 637 AGAAGATTTCCGAATATTAACATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 696  
QY 105 CAGCAGAGAGAAACACTCTCTCGAGATTTACAAAACGTCATCTTACGACAGCTTC 164  
DB 697 GTTAACAGCATCTTCTCTCAGCTGATCTTACCAAGATGATCTCACAGAAAACCCG 756  
QY 165 CACCCACAGTACACCCAGCTTTCTTACGCAATTCAGAGATGATGATGATGATGAT 224  
DB 757 CATCCACAGGTACTGCTGCTTCTTACGATCGTTGATGATGATGATGATGATGAT 816

QY 225 GGAGCAGCAGTCTTCTGATATCTACAGCCACAGAAAGGAGGCTGTGATATGAC 284  
DB 817 GGAAGCAGCAGTCTTCTGATATCTACAGGCTACGAGAGTGAGGTAGCAGTGTAGAT 876  
QY 285 CTGAGTGGGTTGCCAAGACAGCAGTGAATTCGAAGACAGCAGCAGTGAAGACATT 344  
DB 877 TTGACACGCTTCCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 936  
QY 345 GAGAGAGCATCAGATCTCTGATGAGCAGGAGCATTTGTGAGAGAGCTGCTAGAGAGAC 404  
DB 937 ATTGATGGAACAGATCCATTGAGGAGGAGCATCTTGTTCGAAGAAATGCTTTGAAAAGAA 996  
QY 405 CCAATTGACCGGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 464  
DB 997 CCTGCAACAAAATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1056  
QY 465 GCTTTTGCCCAATATGCAATGTCAGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAG 524  
DB 1057 GCATTTGCAAAACATGACATGCTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1116  
QY 525 GTGCTGGAAGAGCAGGAGACCATAGTGAATGATGATGATGATGATGATGATGATGAT 584  
DB 1117 GTGCTGAGCAGGCTGAGAGCTATTTCTTGAAGATGAGGAGAGAGAGAGAGAGAG 1176  
QY 585 GTGATTCGCAATGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 644  
DB 1177 GTTATTTTAAAGCGCAGCTGTGAATGATGATGATGATGATGATGATGATGATGAT 1236  
QY 645 ATGGGAATATGTTTGTGCTCTCTTACCATGAGAGAGAGAGAGAGAGAGAGAGAGAG 704  
DB 1237 ATGGGAATATGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1296  
QY 705 AGAACAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 764  
DB 1297 AGGACTTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1356  
QY 765 CTCATCAAGTATGAAAG 824  
DB 1357 TTAACCATGAG 1416  
QY 825 AAGAAGACCGAG 884  
DB 1417 CATGACATCGGAG 1476  
QY 885 ACCTCAAGAGGTTAAACAATGATTTGGTGAAGAGAGAGAGAGAGAGAGAGAGAGATTC 944  
DB 1477 ACACCTGAGCGTCTCATTAATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGATTC 1536  
QY 945 ATGAGAGAGTCTGTTGACCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1004  
DB 1537 ATGAGAGAGTCTTATTAACCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1596  
QY 1005 AAGTATGAGAGTGTATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1064  
DB 1597 AAGTATGAGAGTGTATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1656  
QY 1065 TTGCTGGTAAATATACATTCATGATGATGATGATGATGATGATGATGATGATGAT 1124  
DB 1657 TTATGGGTAAATATATATTTATGATTTTGAAGGAGAGAGAGAGAGAGAGAGAGAG 1716  
QY 1125 GAAGATTTGAAG 1184  
DB 1717 GAGAGATTTGAAG 1776  
QY 1185 ATGCGGTGCTGCTAATGAG 1244  
DB 1777 ATGCGGTGCTGCTAATGAG 1836  
QY 1245 GCTTCCTTGCCTTTATCTTACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1304  
DB 1837 TCCTCTTCAATTAAG 1896



[illegible]



Db	3379	TCGATCAGCTAAATGTCAGCACTGTGTGACTCCAGCCATGATGATTTCTTCACGGTC	4038
OY	3518	CAGTATGTGTAGCAATTCGTCTTTTGGACTCAGTGGCCAGTCTCAGTCAGCATGAGAGCG	3577
Db	4039	CAGCATTCGTGAGCAATTTGTTGTGTGACTCCATCTCTCAGCTCAGATGAGAACGGTG	4098
OY	3578	CCAGAGGCAATTCGTGCACCATCTGTGAAACCAACCTAGGATGGGCAAGATGAGAGCG	3637
Db	4099	TTCCCTTCAGGCCCTCGGSCAGTCCCTGAAATCCAGTGGGGCATTTGGAAAGACAGAGCAGC	4158
OY	3638	GACCATGATTTAGCACTGATAGTATAGCTTTGGGGTCTATGCACCAATGTCCGAGGGCG	3697
Db	4159	TTTCAGGAGATAGAGAGATCTATAGTCACAACTGCGCCCTGGGTGACACTCTTGAAGCCATCTCT	4218
OY	3698	AGGCTTATATGCTACAGCTACAGTAAATTTCTTCCAAAGCAGAGAGCACTTTCCACAGA	3757
Db	4219	AATCAAGTGTGTTAGCTGTCTCA-----TCGCTGTGTAGCAATGAAAGATTTTCCAGA	4272
OY	3758	TCAGGGGATGCGCGGTGCATTTGATCTGTGTGACAGTGGCGCGTGGGAGCTGGAGCTCATG	3817
Db	4273	GCATATCATTT-----ATGAAAGCAGCTGACGTGTGTGTGAATTTGACTTCTGTG	4323
OY	3818	CTCAGTGTGCTCCCATGATTAATATATACAGACGATCCAGCAGACAGAGAACTTCCGGA	3871
Db	4324	TTTCAGCAGAGCTCCCATGACAACTTCCAAAGCCTTCCAAACCCAAAAGCTGGGA	4377
RESULT 8			
AF070570		1803 bp	Linear
LOCUS	AF070570		
DEFINITION	Homo sapiens clone 24473 mRNA sequence.		
ACCESSION	AF070570		
VERSION	AF070570.1	GI:3387939	
KEYWORDS	FLI CDNA.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens.		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 1803)		
AUTHORS	Anderson,B., Wentland,M.A., Ricafrente,J.Y., Liu,W. and Gibbs,R.A.		
TITLE	A 'double adaptor' method for improved shotgun library construction		
JOURNAL	Anal. Biochem. 236 (1), 107-113 (1996)		
PMID	96307227		
PMID	8619474		
AUTHORS	2 (bases 1 to 1803)		
REFERENCE	Yu,W., Anderson,B., Morley,R.C., Muzny,D.M., Ding,Y., Liu,W.,		
AUTHORS	Ricafrente,J.Y., Wentland,M.A., Lennon,G. and Gibbs,R.A.		
TITLE	Large-scale concatenation cDNA sequencing		
JOURNAL	Genome Res. 7 (4), 353-358 (1997)		
PMID	97264341		
PMID	9110174		
AUTHORS	3 (bases 1 to 1803)		
REFERENCE	Yu,W. and Gibbs,R.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-JUN-1998) Molecular and Human Genetics, Baylor		
PMID	College of Medicine, One Baylor Plaza S930, Houston, TX 77030, USA		
FEATURES	Location/Qualifiers		
source	1..1803		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="T.M.A.G.E. Consortium clone ID 24473"		
	/sex="female"		
	/tissue_type="brain"		
	/clone_lib="INIB"		
	/dex_stage="infant"		
	/note="similar to Homo sapiens mRNA for KIAA0313 with		
	Genbank Accession Number AB002311"		
BASE COUNT	545 a 323 c 346 g	589 t	
ORIGIN			
Query Match	24.1%	Score 1581.4	DB 9; Length 1803;
Best Local Similarity	95.4%	Pred. No. 0;	
Matches 1693; Conservative	0; Mismatches	1; Indels	81; Gaps 3;

QY	4875	CAAGATGATTCACCTCGGGCTGCACTTCTCAATGCCGGAAGGATTTTTTTTAATCTCC	4934
Db	1	CAGATGATTCACCTCGGGCTGCACTTCTCAATGCCGGAAGGATTTTTTTTAATCTCC	60
QY	4935	TTTAGATTTCAATCCAGTCTCTACACTTATCTCATTTGGGATTAATGAGAAAGCTAGCCA	4994
Db	61	TTTAGATTTCAATCCAGTCTCTACACTTATCTCATTTGGGATTAATGAGAAAGCTAGCCA	120
QY	4995	TTGGAATCTTTGGGGGCTTTAACCACCAAGAGACAAAGAAAAACAATGAATCCTTT	5054
Db	121	TTGGAATCTTTGGGGGCTTTAACCACCAAGAGACAAAGAAAAACAATGAATCCTTT	180
QY	5055	GAGTACAGTGTCTGCTCCACTCTGTTTCAATGTCCTCCTTT-AAAAAATAATGAGTT	5112
Db	181	GAGTACAGTGTCTGCTCCACTCTGTTTCAATGTCCTCCTTTAAAAAATAATGAGTT	240
QY	5113	TAAAGATTTTCTCAGAGATAAATATATATCCATTATTAATGATTCAGATTAATTTTAA	5172
Db	241	TAAAGATTTTCTCAGAGATAAATATATATCCATTATTAATGATTCAGATTAATTTTAA	300
QY	5173	CCTTAAGTAGGGTTGCCAGGCTGCTTTCGAAAAACAATATGCGGACAGGGGTGGGC	5232
Db	301	CCTTAAGTAGGGTTGCCAGGCTGCTTTCGAAAAACAATATGCGGACAGGGGTGGGC	360
QY	5233	CACACCAAGAGAGGGAAGACCTGCTGTGACCCCTGGCTCCCATGTCTCTGCTCT	5292
Db	361	CACACCAAGAGAGGGAAGACCTGCTGTGACCCCTGGCTCCCATGTCTCTGCTCT	420
QY	5293	CACCCGCGAAGTGGCCCTATCTCTGGAAGTATGAATGTTAGCCAAATTAATACCAAGACCC	5352
Db	421	CACCCGCGAAGTGGCCCTATCTCTGGAAGTATGAATGTTAGCCAAATTAATACCAAGACCC	480
QY	5353	TCAATCTGCTCTTCCCGAGTGAATGGGGTCTCTCTAATACCTGTTGCACATGAGCCAG	5412
Db	481	TCAATCTGCTCTTCCCGAGTGAATGGGGTCTCTCTAATACCTGTTGCACATGAGCCAG	540
QY	5413	GGAGGAACTAGAGACCTTGTCTGTCTGAGCCTTAATGAGGAGAGAGCGTGTCAATG	5472
Db	541	GGAGGAACTAGAGACCTTGTGTCTGTCTGAGCCTTAATGAGGAGAGAGCGTGTCAATG	600
QY	5473	GCGGATGTGCTCTCTCATTTGAGATGAGTGGCAAAACCCATTATTAAGTATATATCTT	5532
Db	601	GCGGATGTGCTCTCTCATTTGAGATGAGTGGCAAAACCCATTATTAAGTATATATCTT	660
QY	5533	TGATTTTGTAAATTTAGAGGTGTAGGTTTG-TTTTTGTTTTTGTTTTTTTTAA	5591
Db	661	TGATTTTGTAAATTTAGAGGTGTAGGTTTGTTTTTTTTTTTTTTTTTTTTTAA	720
QY	5592	GAACATTTTAACTGATATGCAATTCGAGTGAAGAGAGCTTGGGATGTTGAGCTAATGC	5651
Db	721	GAACATTTTAACTGATATGCAATTCGAGTGAAGAGAGCTTGGGATGTTGAGCTAATGC	780
QY	5652	CAGCTGTTTAACTGCTCTTCAAGACAGCTC-----5684	
Db	781	CAGCTGTTTAACTGCTCTTCAAGACAGCTCCTTTATTAATGTCATTAGGGAATA	840
QY	5685	-----CCTTAATG	5693
Db	841	AACAAGCCTTTAAAGCTGATTAAGAAAGATCAAAAACTGGTTAACAATGCAAGCTTTAATG	900
QY	5694	AATTGCAATTTGGGAATAAACAAGCCCTTTAAACGTGATATAAAGATCAAAAACTGGTTAG	5753
Db	901	AATTGCAATTTGGGAATAAACAAGCCCTTTAAACGTGATATAAAGATCAAAAACTGGTTAG	960
QY	5754	ACATGCCAGCCTTTGCAAGGAGAGTATAGTACCAAGAGATTAACCTCAAGTGGCTTATG	5813
Db	961	ACATGCCAGCCTTTGCAAGGAGAGTATAGTACCAAGAGATTAACCTCAAGTGGCTTATG	1020
QY	5814	GACGCTGATATAGAGAAGCCTTAAGTGTAGCAACCATCTGCTCACAGCTGATTAAC	5873
Db	1021	GACGCTGATATAGAGAAGCCTTAAGTGTAGCAACCATCTGCTCACAGCTGATTAAC	1080

5874 CTATATGACTGTAATGACCCCTCAGCTATTTTGTGTGTTTGACAGACTCCGGA 5933  
1081 CTTAATATGACTGTAATGACCCCTCAGCTATTTTGTGTGTTTGACAGACTCCGGA 1140  
5934 AAGTGAAGGCTCCCAATCTGAGTACTCAATGTGAGAACCTGCTTGGATTT 5993  
1141 AAGTGAAGGCTCCCAATCTGAGTACTCAATGTGAGAACCTGCTTGGATTT 1200  
5994 TTTTTCATTAATTAATGATGATCATATTGATGATGATGATGATGATGATGAT 6053  
1201 TTTTTCATTAATTAATGATGATCATATTGATGATGATGATGATGATGATGAT 1260  
6054 TTTTAAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6113  
1261 TTTTAAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
6114 TAAATCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 6173  
1321 TAAATCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380  
6174 TGTAAATTTTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6233  
1381 TGTAAATTTTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
6234 TGTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6293  
1441 TGTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
6294 ATTAAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6353  
1501 ATTAAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
6354 GAATGATCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6413  
1561 GAATGATCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620  
6414 TTTTGAAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6473  
1621 TTTTGAAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
6474 TATTTGTTATTTTGTGTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6533  
1681 TATTTGTTATTTTGTGTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740  
6534 GTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6568  
1741 GTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1775

RESULT 9  
AF478468 4806 bp mRNA linear PRI 25-FEB-2002  
LOCUS AF478468  
DEFINITION Homo sapiens Rapi guanine nucleotide-exchange factor PDZ-GEF2A  
ACCESSION AF478468  
VERSION AF478468.1 GI:18874697  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 4806)  
AUTHORS Kuijperl,H.B., de Rooij,J., Rehmann,H., van Triest,M.,  
Wittinghofer,A., Bos,J.L. and Zwartkruis,F.J.T.  
TITLE Characterization of the PDZ-GEFs, a family of guanine nucleotide  
exchange factors specific for Rapi and Rap2  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4806)  
AUTHORS Kuijperl,H.B., Zwartkruis,F.J.T. and Bos,J.L.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-2002) Physiological Chemistry, University Medical  
Centre, Universiteitsweg 100, Utrecht 3584 CG, The Netherlands  
FEATURES Location/Qualifiers

Source 1. 4806  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
CDS  
1. 4806  
/note="PDZ-GEF2"  
/codon\_start=1  
/product="Rapi guanine nucleotide-exchange factor  
PDZ-GEF2A"  
/protein\_id="AA179915.1"  
/db\_xref="GI:18874698"  
/translation="MNSVDPGCAQALRRKRPPTEDLNTIYSYLGMEILNLRH  
OLRMSARPERYSGNOVLESETIARCVTLISGSVLYKVSIVLPSCFGQFGK  
RGCCIVARERYSKNOVLESETIARCVTLISGSVLYKVSIVLPSCFGQFGK  
NSYSLPADLILKMLTENPHQYVHVSOSGCSIASDSGSSLSDDIYQATSESGV  
DLTRIPESVSEDEDEIDTRDPLQGDLYRELEKPERAKPTDIDLELFPH  
OLPAFANMTMSVREELCSVLEVEVAGAILLEDGELSWVILNGVTEISHPDK  
VENLEMGSEFIPPTLKOYHNGIVRTVDDQCPVCAQDDYMWILNHRNTHKVE  
EGEIVMVEHELEDRSGTRKGIHIVKAPERLIMHLEHSIDVPVTEIDFLYRTF  
LESPLDVGILKLEWFKIDSLRDYTRIVLWNNHNFDEPDMPREFLEFEKLEED  
TKMNHLEPLNITACAAKAKRWVLOKASRESPLQFSLNGSEKFGFIEVEGPGE  
AADSLKRGDDIMEYNGONFENTPMKAVELIRNTHLALIVKNIIFVKELEPERED  
EKSGVPHIPKIAKKSNNHSIQHPGDIEDTSOKGSKKRYKANTASGGRKIRIILDK  
TRFSILPEKLESDGLOSODSDIVTGRCHSLALMTPELTSSSPDLQPTSL  
DFNSDIPDQVIRVFKVDOOCYIIISKDTAKVEVFAHVEHGLGASDYSLCEV  
SVTEPGVAKORRLEPDQFSLADRIOLNMGVLYKNNMETELCSDEDOELVKEQSLSM  
LOLSIEVATLOSMRDFLEFNIEPTVIDLFLNKTGNTHLKREFIDVNOETFW  
ASEILTEANOLKRMKIIKHFIKIALHGECKENFMSFALISGLASVARGTWEK  
PSYKIRHLODQIDIEDPSRMNAKRNILSSOSMOPPIIPEPVYKRMPTFLHEDNSK  
VGLVNEFKLRNLSKEINOVNMTSANDPMHMRORSLSOGSTNSMIDVGGANHK  
BARSSLNANLKLIEDAOMARKVOYLSLDEVEDEEKFQWMSLOMERPAETLNLS  
EKRSKSEMSVPMRSGQTKALHQPHVSOVLOVPAVNLHPIRKKGOTKOPALN  
TSLPQKVLGTEIEISGRKTEDTISVASSPSAPSGSPHKGTYLIPSKDNLNLS  
DSHSEISRSRSIVNCSVDMSALDOERCSOALVPESTGALLETTHASGIGDS  
OHGPGMTLPSLILKLAIVSSVNEELISOHIIIEADSRGSMTESSSHDNFOS  
LPNPSMPLNSYRTHDIPDAVPEPDSRYGCSKSCSPGCGCKGCELEKSWTS  
SSLSDEYENGTGVARVLESTPABSEGLPKDPTDYRTYTVSSTEKGLIYVCVNS  
PKKDRYRPPPTPGVGLISLADIEKSPHHLKPPDVSVAVQSKMHNLSLPPA  
SLSSNLVACVPSKIVTOPORHNLDPFHRLKLDVDVDADEADENQVSAV"

BASE COUNT 1508 a 953 c 1081 g 1264 t  
ORIGIN  
Query Match 24.0%; Score 1577.6; DB 9; Length 4806;  
Best Local Similarity 65.5%; Pred. No. 0;  
Matches 2508; Conservative 0; Mismatches 1229; Indels 93; Gaps 10;  
45 AGCATATGCTTTCTTCATTTATGAAACCACTAGCAATCCAGTAACTAGGATTTAGGC 104  
418 AGAAGATTTGGAAATTAATTAAGAGAGAGCCAAACATTTCATGATGATGGAG 477  
105 CAGCAGAGAAACATCACTCTCGCAGATTTCAAAAACGTCATCTTACGACTCTC 164  
478 GTTAACAGCTATCTTCTCTCCAGCTGATCTTACCAACATGCTATCCACAGAAACCT 537  
165 CACCCACAGGTGACCCAGCTTTCTTAGCCATGAGATGATGATGATGATGATGATGAT 224  
538 CATCCACAGGTGACTATGCTCTTAGCTAGCTGCTGTTGATGATGATGATGATGATGAT 597  
225 GGGAGCAGAGCTCTTCTGATATCTACAGGCCACAGAAAGCAGGCTGGTATATGAC 284  
598 GGAAGCAGAGCTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
285 CTGAGTGGTGGCCAGAAACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 344  
658 TTGACACGCTTCCAGAGAGACCTGTTGATCTTGAGATGATGATGATGATGATGATGAT 717  
345 GAGAGACATCAAGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 404  
718 ATTGATGAAACGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 777  
405 CCAATGACCGAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 464  
778 CCTGACACAAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837



D 2992 AATATCTTAGTAGTCAAAAGTATGCAAGCCTCCAAATTATTCCTACTCTTCCTGTTGTCAG 3051  
Q 2685 AAGATCTCACTCTTCTTACGAAAGAAATGACTCAAAATTAACGGCGCTGGTCAATTTT 2744  
D 3052 AAGATATGACATTTCTTACATGAAGAAATGACTCCAAATAGATGGTTAGTAACTTT 3111  
Q 2745 GAGAGCTAGAGTATGATGGAAAAAAGAAATGTCAGCTTGGCCGAATGGCTTAGTGAAC 2804  
D 3112 GAGAAAGTAAAGATGATTTCCAAAGAAATCCGCCAAGTTTTCGATGACTTTCGCTAAC 3171  
Q 2805 ATGAGACCTGCCCCATGATTCAGAGACTCGAAGAAAGAAATGGCGAGTTTGGGCTCTC 2864  
D 3172 ATGAGACCTGATGATGCTT-----TCACAGAGAGTCACTG 3207  
Q 2865 AGCAGGGTATGATCAATATGCAACAGTCTTGATGTTCTCTCAGAGAGTGGTCAATAAAG 2924  
D 3208 AGTCMAAGAAAGCAAAATTCMAACATCTGATGT--TCAGGAGGTGCTCACAAAAA 3264  
Q 2925 CGGGTAGGTGTTGTTCTTTCGATGCGCAAAAGCTTTATGAAATGGCCAAATGGT 2984  
D 3265 AGGCGACCGCGCACTCTTCTGCTTAATGCGCAAGAGCTTATATGAGATGCCCAATGGCA 3324  
Q 2985 CGAAAGTGAAGCAGTACTTTCGAATTTGGAGCTAGAAATGACGAGAGAGTCTTCA 3044  
D 3325 AGGAAGGTGAAGAGTATCTTTCCAGTCTCGATGATGAGACAGATGAGAGAAAGTCTCAG 3384  
Q 3045 ACATATCTCTGCGAGTGAAGCCGACCAACACATTCCTTAAGATTCCTGGTGACAAA 3104  
D 3385 ATGATGATATTAAGAGTGGAGCGCTGATATGATGATCTTGACCAAGAAATTAAGTGA 3444  
Q 3105 AAGCCTTCAAA--ATCCGAGACCTCTCCAGTACTCCAAAGGCGAGGCTCAACACAGAAA 3161  
D 3445 AGATACGCAAAATGATCTGAAATGTCTCCAGTGTGAGAGTCACTGGCGCAAAAGACT 3504  
Q 3162 GCTCAGTCCCTGCGACAGAGCCGACAGCAGCAGCAGCAGTAAATCAACAGAGGA 3221  
D 3505 AAGCCCACTTTCATCAACC-----CACAGAGTAAGCCAGGTG 3543  
Q 3222 CTACAGTTTCCGCGCTGCTTATCTTTCACGAGAAAGAGTCCGCTAAAGATCTC 3281  
D 3544 CTTCAGGTGCGCACTGTTAATTTGCAACCCCTTCAGAGAAAGGCAAAAGAGACCT 3603  
Q 3282 CCACCTTTTGGCATTAATCTTCACAGCTTTTAAAAAAATTTCTTCTGCTGAAGAA 3341  
D 3604 GCACTGATACAGTTTACCTCAGAAAGTTTACAGACAA-----CTGAGAA 3651  
Q 3342 CGAAGTTTGGAGCTCAGAGAAACAGGCTGAAATATCAATATGATGATCTTCCAG 3401  
D 3652 ATAA-----GTGTAAGAACCATACAGAGACATATTTCTGTCCTCATCTTTA 3702  
Q 3402 CTTCCTCTCTCTCTCTCTCTCCACAGAGTCTCCAGGAAAGGCTTATCTTGGCTCC 3461  
D 3703 CATCTCTCTCTCTCTCTCTCTCCAGAGCTCCCTCACAAGGTTTACACTTATTC 3762  
Q 3462 AGTGTACTGTGATTAATTTTTCAGATTTCTGTCAGAGTAAATTTCTTCAAGTCCAGT 3521  
D 3763 TCAGCTAAATCTGACAACTGTCTGACATCCAGCATAGTGAATTTCTTCAAGCTCAGC 3822  
Q 3522 ATTTGTAGCATTTGCTTTTGTGACTCAGTGCAGTCTCAGTGCAGATGAGAGCGCCAG 3581  
D 3823 ATCTGTAGCATTTGCTTTGTTGACTCAGTGTCTCAGTCTCAGAGATGAAGCGGTGTTCC 3882  
Q 3582 AGGCAATTTGTCAGATGCTGGAAGAAACCTAGGAGTGGCAGATGAGAGCGGAGC 3641  
D 3883 TCTCAGAGCCCTGGCAGTCCCTGAAATCCACTGGGCACTTGGAAAAACAGACAGAGCTTCA 3942  
Q 3642 ATGATGAACCTGATCAATATAGCTTGGGCTCTATGACCAATGTCGAGGCGGAGGC 3701  
D 3943 GGGATAGAGATCATATGCAACATGAGCCCTGGGAGACACTTGAAGCCATCTCTATAC 4002  
Q 3702 TTATATGCTACAGCTACAGTATTTCTTCTCAGACAGAGAGAACTTTCCAGAGATCAG 3761  
D 4003 AAGGTATTAGCTGTCTCA-----TCGTCTGTGAGCAATGAAGATTTTCTCAAGAGAT 4056

Q 3762 GGGATCGGCGCTGCACTGATGCTGACAGTGGCGCGGAGCTGGACATGATCTCA 3821  
D 4057 ATCATT-----ATACAGCAGCTGACAGTGGCTGGAGAGTTGACATTCGATTTCA 4107  
Q 3822 AGTGCTCCCATGATTAATATACAGAGATCCAGCACAGAGAGTGGGA 3871  
D 4108 AGCAGCTCCCATGACAACTCCAAAGCTTCCAAACCCAAAAAGCTGGGA 4157

RESULT 10  
AF478469 4176 bp mRNA linear PRI 25-FEB-2002  
LOCUS AF478469  
DEFINITION Homo sapiens Rap1 guanine nucleotide-exchange factor PDZ-GEF2B  
ACCESSION AF478469  
VERSION AF478469.1 GI:18874699  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 4176)  
Kuiperij,H.B., de Rooij,J., Reijman,H., van Triest,M.,  
Wittinghofer,A., Bos,J.L. and Zwartkruis,F.J.T.  
TITLE Characterization of the PDZ-GEFs, a family of guanine nucleotide  
exchange factors specific for Rap1 and Rap2  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 4176)  
Kuiperij,H.B., Zwartkruis,F.J.T. and Bos,J.L.  
TITLES Direct Submission  
JOURNAL Submitted (28-JAN-2002) Physiological Chemistry, University Medical  
Centre, Universiteitsweg 100, Utrecht 3584 CG, The Netherlands  
FEATURES  
source  
CDS  
1..4176  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..4176  
/note="PDZ-GEF2"  
/codon\_start=1  
/product="Rap1 guanine nucleotide-exchange factor  
PDZ-GEF2B"  
/protein\_id="AA179916.1"  
/db\_xref="GI:18874700"  
/translation="MNSPYDPGARGALRKKPPERTPEDLNTIYSLHGMETLSNLRH  
OLRMSARARERYSGNOYLFCESTIARWYLLSGVLYGSMVLPSCFGKQFGK  
RGDCIVLEPSMIVENAKKNEDSLREIPIAROSRRFRINVKGRQITTDVEY  
NYSLSLPADLTLMHLEPNHPDYTHYSSQSCSTASQSGSSLSDIQATPESPGV  
DLRLPEGVDEDEDEDEDEDETRDPLDGRDLRECELEKREPADTDDIDQLFPMH  
QLPAFANMTMSVRELCSYVMEVEVOAGAILLEDGELDSVYLLNGTVELSHPDGK  
VENLEFNGNSFGITPLDKQYMHGIVRTKWDQCFVCIADQDWRLINHEVNTHKVE  
EGEIVVHEHRELDNSGTRGKHIVIKAPERELIMHLEHSLVDPYTEDFLVYRT  
LESPLDQVFKILLEMFKIDSLRDKVRIYALVWLNHNHPDEGPATRLPEEKNLE  
TKNGHLRLNLNTACAKAKAROVYLOKASRESPILOFSLNGSEKFGYFVCEVGESE  
AADSGLKRDQIMEVNGQENETFMKAEILRNTHLALVYKNIYFKELPRTEDQ  
EKSGVPHIKPIAKKSNRHSIQHVPGEDLOEOTQESKSKVKVKNATSGSRNKRITLIDK  
TFEFSILPKLFEFDGSLQSDSIVGTRCHSLAIMPPTLSSSSDLDLQPTSM  
DPSNPDIIDQVIRVEKVDOSQCYIILISDRCAVEVFAHVEFLTASDYSLECV  
SVTPREGVIRKORLPODFSLADRIOLNGYTYLKNMMEETLCSDDAOLEVSEQLSM  
LQSLTEVATQLSMDFDLFRNIEPEVYDLEFKNSKGTGNHLKRPDIYVQETFMV  
ASFTLLEAQLDQIDFDPSSNRMAKRYNLISSQSMOPPIPLFVYVAKDMTFLEHNDK  
PSKYEKHLDQIDFDPSSNRMAKRYNLISSQSMOPPIPLFVYVAKDMTFLEHNDK  
VGLVNFERKLMDISKEIRGVVMTSANDPAMFQORSLSQSTNSMLDYOGGANK  
BARRSKSLNAKKLYEDQAKRKQYLSLDEVTDEKFOMSLOMPAYGTLTNLS  
EKRSASSMSKPVPMRSAGCTTCAHLHOBHRVSOYLOVAVVNLHPIRKGQTKDPA  
TSLPOKVLGTTEBEISGKKTEDPTIYASLSHSPASPSGPHKYSILSHSSKISG  
QSCPGTGGATLQKTLQITRTAKRTDSTEFKTEENRDRPTSCENTTRKRMSPFRLK  
ERMLSRERLVNSQKEDTDHNOATESECKYDVGSNIKDEKGSALFNSNSQGSNPLNC  
FYTRFSKRRKTL"

BASE COUNT 1378 a 776 c 944 g 1078 t

Query Match

22.9%; Score 1502.6; DB 9; Length 4176;

Best Local Similarity 68.7%; Pred. No. 0;  
Matches 2158; Conservative 0; Mismatches 949; Indels 36; Gaps 5;

OY 45 AGCATATGTTTCTTCTATTTGAACCACTAGCATCCAGCTAACCATGAGATTAGGCG 104  
DB 418 AGAAGATTTTCGAAAAATCTAATAAGGAGCGCCAAACCATTAATGATGTCGAG 477  
OY 105 CAGCAGAGAAACACTCCTCTCTGAGATTTGACAAAACCTGATCTCTGCACTGAC 164  
DB 478 GTTAAACGCTATTTCTTCTTCCAGCTGATCTTACCAAGATGATCTCTCAGAAAAACCT 537  
OY 165 CACCACAGGTGACCCAGCTTTCTTCTAGCCATTGAGATGATGATGATGATGATG 224  
DB 538 CATTCACAGGTGACGATGTCCTTCTAGCACTGTCGTTGATGATGATGATGATGATG 597  
OY 225 GGGAGCAGAGCTTTTCTGATATCTACAGCCACAGAAAGCAGGCTGGTATGAGAC 284  
DB 598 GGAAGCAGAGCTTTATCTGATATCTACAGGCTACGAGAGTGAGGTAGAGATGATGAT 657  
OY 285 CTGAGTGGGTTGCGAAGAACAGAGTGGATTCGAAAGCAGAGCAGATGAGAGACATT 344  
DB 658 TTGACACGCTTCTCAGAAAGAGCTGTTGATTTCTGAGATGACGAAAGAGATGAGAG 717  
OY 345 GAGAGAGCATCAGATCCTCTGATGAGCAGGACATTGTGAGAGCTGCCTAGAGAGAC 404  
DB 718 ATTGATCGAAGACAGATCCATTCGAGGGCGAGATCTTGTGAGAAATGTCCTGAAAAAGAA 777  
OY 405 CCAATTGACCGGACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 464  
DB 778 CCGAGACAAAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 837  
OY 465 GCTTTTGGCCAAATGACAAATGTCAGTGGAGCGAGACCTGTGCTGATGATGATGATG 524  
DB 838 GCAATTTGCAAACTGACATGTCGTGTAAGAGAGAACTCTGCTCAGTGTGATGATGATG 897  
OY 525 TTGTTGGAAAAAGCAGAGGACCATAGTGTAAATGATGTGAAGAGCTGACCTCTGCTCA 584  
DB 898 GTGCTGAGACAGGCTGAGCTGATTTATTTCTGAAATGAGGCAAGAGCTGATGATGATG 957  
OY 585 GTGATTTCTCAATGATCTGTGGAAGTACTATCCAGATGGAAGAAACAGAAATACGTG 644  
DB 958 GTTATTTTAAAGCGCACTGTGGAATCAGTCATGATGATGGAAGAAATGGAATTTGTTT 1017  
OY 645 ATGGGAAATAGTTTGTGTCCTCTACCATGAGCAAAAGATACATGAAGAGAGTATG 704  
DB 1018 ATGGGAAATAGTTTGTGATTTACTCCACTCTGTGATAGCAGTACATGCAATTTGTC 1077  
OY 705 AGAACAAGGTGATGATGTCAGTTGTCTGATAGCCGACGACAGATTTACTGCCATTT 764  
DB 1078 AGGACTAAAGTATGATGATTTGTGCTGATAGCCGACGACAGATTTATGAGCAATT 1137  
OY 765 CTCGAATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 824  
DB 1138 TTAAACCATGTGGAAGAAATACCATTAAGAGAGAGAGGAAAGAAATTTGATGATG 1197  
OY 825 AAAGAACACCGAAGACTGTGCAACTGGAACAAGAAAGGACACATGTCATCAAGGAT 884  
DB 1198 CATGAGACATCGGAACTAGACCGGAGTGGACAGCAGAAAGACATGTCATCAAGGCA 1257  
OY 885 ACCTCAAGAAAGTTAAACAATGCAATTTGTGGAAGAGCAATTCAGTATGATCAACATT 944  
DB 1258 ACACCTGAGCGTCTCAATATGCAATTTAATGAGAAACATTCATCGTATCCAACTTAT 1317  
OY 945 ATGAGAGACTTCTGTGACCTATAGGACTTTCTTTAGCCCATGGAAGTGGGCAAA 1004  
DB 1318 ATAGAAGATTTTCTATTAATCTTACAGACATTTCTGGAAGCTCTTGTGGATC 1377  
OY 1005 AAGTTATTTGAGTGTGTTAATGACCCGAGCCTGAGGATAGGTTACAGCGGATGATTA 1064  
DB 1378 AAACCTTTGGAATGTTTAAGATCGACAGCTTAAGAGATAGGTGACACGATTTGATTA 1437  
OY 1065 TTGTGGTAATATATCACTTCAATGACTTGAAGAGATCTTCAATGATGATGATTTTA 1124  
DB 1124

DB 1438 TTATGGTAATATATCATTTTAAATGATTTTGAAGTGTACCCCTGATGATGATTTCTA 1497  
OY 1125 GAAGATTTGAAACATCTGTGAAAGAGAGAAATGGGTGACACTTAAGGCTGTTGAAT 1184  
DB 1498 GAGGAATTTGAAAAATCTGGAAGATACAAAGATTAAGTGTATCTCCGGTTATGGAAT 1557  
OY 1185 ATGCGTGTGCTCTTAAGCAAAAGATGATGACGTTATCAAAACCATCCGAGAA 1244  
DB 1558 ATTGCTGTGCTCAAAAGCTTAAGTGACAGAGTGTGTGCTGCAAAAGGCTTCCCGAG 1617  
OY 1245 GCTCTTTGCTTTTATCTTACTTGGAGGCTCTGAGAAAGGATTTGGAATCTTTGTGAC 1304  
DB 1618 TCCCTCTCAATTTCAAGCCTTAATGAGGAGTGAAGAGGATTTGATTTTGTGGA 1677  
OY 1305 AGTGTGATTCAGTGTGCAAAAGCACTGAACGCTTGAACGGGGGATCAATATTA 1364  
DB 1678 GGAATGAACTGCTGTGAGCGAAGCTGCTGATTCAGCACTGAAAGCTGTGATGATGAT 1737  
OY 1365 GAAGTAAATGGCCAAACCTTGAACATTCAGCTGCAAAAGCTTATGGAATTTCTAGA 1424  
DB 1738 GAAGTAAATGACAAACCTTTGAAATATTACATTTATGAAACCGTTGAAATTTTGAGG 1797  
OY 1425 AATACACACATTTATCTATCACTGTGAAACCAATTTATTTATTTAAGAACTTTCTA 1484  
DB 1798 AATATATCTATCTGCACTTACTGTGAAGACCAACATTTTGTGTCAAAGATTTACT 1857  
OY 1485 ACAAGATTCGACAAAGAAAGAAATGTGCCCCCACCCTTCTTAATTTGTTGACATT 1544  
DB 1858 TTTAG--GACTGAACAAAGAAATCTGGGCTTCCTATTTCCCAAAAT--TGCAAAA 1911  
OY 1545 AAAAAGCCAGTGTCTACTTCCAGATCTTGTGATGATGATGATGATGATGATGATG 1604  
DB 1912 AAAAAAGTATGCTCATTTCTATTCAGATGTCCAGAGAGATTTGACAGACATCAAG 1971  
OY 1605 CTTGAAAAAGTGAACAAAAAAGTAAAGCCACACTGTGGAGAGAAAGCAAGCTGAAA 1664  
DB 1972 GAGAAAGAAAGTAAAGAAAGTTAAAGCAAAATACGCTTCAAGTGAAGAAACAAATCAG 2031  
OY 1665 AATATCTGACAAAGACTGGATCAGTATGTTGTCACAGAAACCATGATGATGATG 1724  
DB 2032 AATATTTGATTAACACAGCATTTAGTATCTTCCCTCAAAAGCTATTTAGTATGAGG 2091  
OY 1725 ATGTCAGTCTCAAGATGACAGATAGTATGATGATGATGATGATGATGATGATG 1784  
DB 2092 CTAAAGCAATACAAATGACAGATTTGTGGAAAGAGCAGTGTAGGCTATGCTGCT 2151  
OY 1785 GCATTCCTGTGAGTGAACCTTATCATCCAGTAAATCTGATTTATTTGACAGTACAT 1844  
DB 2152 ATATGCCCCATCCCTGAAACACTCTCATTCACAGCCCTGATGCTCCGAGCCTACCA 2211  
OY 1845 CGCATTTTACATTCAGTGTCTGACTGACTGCTGCTGCTGCTGCTGCTGCTGCTG 1904  
DB 2212 AGTATGTTGATTTTTCATCTCTGATGATGATGATGATGATGATGATGATGATG 2271  
OY 1905 GCTGATCAGCAAGCCGCTCATCATGATGATGATGATGATGATGATGATGATGATG 1964  
DB 2272 GTGATTCAGAAAGTGTCTATCATTTATCATGATGATGATGATGATGATGATGATG 2331  
OY 1965 ATTCAAGCTATCAAGAGATTTGCTTACTGCCACCCCGGATCAATATTTCAATGATG 2024  
DB 2332 TTTCATGCTGCTCATGAATTTGTTGACCGGCTGATCCGACACATATTTCTCTGGA 2391  
OY 2025 GTCTCTGTACACTGAGGAGTAAATCAACAAGAAAGACTTCCAGATGCTTTCCAA 2084  
DB 2392 GTTTCGTGTAATCTTGAAGGTGTCAATTAACAGAGAGACTTCAGATGCTTCCAA 2451  
OY 2085 CTGACAGACATTAACAGTGAAGTATTTATCTGAAAAACAAGAGAAAGAA 2144  
DB 2452 TTAGCTGATGAATTAACATCAATGGAAGTATTTACTTAATAAATAACATGGAACAGAA 2511  
OY 2145 ACTCTTTGTTCAAGTGAAGAGTGTCAAGAGCTTGTGAGAGAGATCAAAATTTCCCTCT 2204  
DB 2512 ACCTTATGTCAGATGAAGATGCTCAAGAACTAGTTAAGGAAAGCAGTATCATGCTG 2571

```

QY 2205 CAGCTCAGCAGCTGTGGAAGTTGCAACAGCCTCTGATGGGAAATTTTGAACCTCTTCCG 2264
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2572 CAGCTCAGTACCATTTAGAGTGGCCACCAGCTGTCAATGAGGAGCTGTTGTTCCG 2631
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2265 AACATTGAACCTTCTGATATATAGATGATTTTAACTGATCAAAACCAGCTGT 2224
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2632 AATATTGAACCGACTGATGATGATGAGCCTTTTAAATTAATTCACAAACAGGAAAT 2691
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2325 GCCAAGCTGAAGAGATTTTGAAGAGTCATTAACAGGAAACATTTTGGTACATCTGAA 2384
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2692 ACCTCATTTGAAGAGTTTGGAGCATTTGTAACCAAGAGACATTTGGTGGCTCAGAA 2751
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2385 ATTCTCAGAAACAAACAGCTGAGAGAGATGATCTTTAAGCATTTTCAATCAGATA 2444
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2752 ATTTTAAGTCAAGCAATTCAGCTCAAAACGAAATGAAATTTTAAAGCATTTTAAAT 2811
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2445 GCATGCTAGTGTGGGAAATTCAGAAATTTTAACTCAATGTTTCAATCATCAGTGGCTA 2504
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2812 GCATCTATTGTGAGATATTAAGAACTTAATTCATGTTTGCATTAATTAAGTGGCTTG 2871
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2505 AACCTGGCACCAGTGGCAAGACTGCGAAGCAGCTGGGAGAACTTCCCAATTAATACGAA 2564
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2872 AACCTGGCATCTGTAGCAAGACTGAGAGAACTTGGGAAAGTTACCAAGCAATACGAG 2931
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2565 AAATCTATTTCAAGATCTCCCAAGACCTTTTGATCTTCCGAAACATGGCAAAATATCCT 2624
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2932 AAACATCTTAAGATCTCAAGACATTTTGTATCCATCTAGAAACATGGCAAGATAGA 2991
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2625 AATGTTCTCATAGTCAAAATCTACACCTCCATATCCCTATCCCTAGTATATCAAA 2684
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2992 AATTTTCTTGTAGTCAAAAGTATGACGCTCCATTTATTCACCTCTTCCCTGTTGTCAAG 3051
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2685 AAGCATCTCACTCTCTTCAAGAAAGAAATGACTCAAAAGTACGCGGCTGTCAATTTT 2744
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3052 AAAGATATGACATTTTACATGAAAGAAATGACTCCAAAGATGCTTGTAGTAACTTT 3111
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2745 GAGAACTAAGATGATTTGCAAAAGAAATTCGTACCTTGGCCAAATGGCTTCAGTCAAC 2804
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3112 GAGAACTAAGAAATGATTTTCAAGAAATCCGCCAAGTTGTCAATGACTCTGTCTAAC 3171
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2805 ATGACCCCTGCTCATGTTTCAGGACTCGAAGAAATGGCGAGTTTGGGCTCTCTC 2864
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3172 ATGACCCAGCTATGATGTT-----TCGAGAGAGGTCTACTG 3207
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2865 AGCCAGGCTACTACAAATGCAACAGCTCTGATGTTGCTCAGACAGTGTGTCATAAAAAG 2924
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3208 AGTCAAAGGAAGCACAAATTCAAACATGCTGATGT---TCAGGAGGTGCTCACAAAAAA 3264
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2925 CGGCTAGCTGCTAGTCTCTTCTCAATGCGCAAAAGCTTTATGAAGTGGCCAAATGGCT 2984
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3265 AGGCGACGCGCCAGCTCTCTGCTTAATGCCAAAGAACTATATGAGATGCCCAATGGCA 3324
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2985 CGAAAGTGAAGAGTACCTTTCCAAATTTGGAGCTAGAAATGAGAGAGAGTCTTCTAG 3044
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3325 AGGAAGTGAAGAGTATCTTCTCAGTCTGATGATGAGAGATGAGAGAAAGTCTCAG 3384
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3045 ACATTAATCTCTGAGTGTGAGCCAGCAACCAACATTTGCTTAAGATCTCGGTGACAAA 3104
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3385 ATGATGTCATTAACAGTGGAGACCTGCATATGCTTGCACCAAGAAATTTAAGTAGAAA 3444
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3105 AAGCTGTGCA--ATCGAGACCTCTTCACATCTCCAAAGGAGGCTCCACAGAAA 3161
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3445 AGATTCAGCCAAATATCTGGAATGTCTCAGTGGCTATGAGGTCAAGCTGGCCAAACAAC 3504
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3162 GCTCAGTCCCTGCCACAGCCCA 3184
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3505 AAAGCCACCTTGATCAACCCCA 3527
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 11
AC124358/c
LOCUS AC124358 181796 bp DNA linear HTG 04-JUL-2002

```

```

DEFINITION Mus musculus chromosome UNK clone RP24-491B24, WORKING DRAFT
SEQUENCE, 10 unordered pieces.
AC124358
AC124358.2 GI:21693950
VERSION 1
AUTHORS McPherson,J.D. and Waterston,R.H.
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 181796)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
2 (bases 1 to 181796)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 181796)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Jul 4, 2002 this sequence version replaced gi:21426479.
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M.BB0491B24
----- Summary Statistics -----
Sequencing vector: M13: 0%
Sequencing vector: plasmid: 100%
Chemistry: Dye-primer ET: 0% of reads
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990319
Consensus quality: 175933 bases at least Q40
Consensus quality: 177034 bases at least Q30
Consensus quality: 177034 bases at least Q20
Insert size: 181582; sum-of-ctrls
Quality coverage: 10.97 in Q20 bases; agarose-fp
Quality coverage: 9.38 in Q20 bases; sum-of-ctrls
-----

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 5779: contig of 5779 bp in length
* 5780 5879: gap of unknown length
* 5880 21352: contig of 15473 bp in length
* 21353 21452: gap of unknown length
* 21453 44568: contig of 23116 bp in length
* 44569 44668: gap of unknown length
* 44669 77089: contig of 32421 bp in length
* 77090 77189: gap of unknown length
* 77190 100595: contig of 23406 bp in length
* 100596 100695: gap of unknown length
* 100696 125457: contig of 24762 bp in length
* 125458 125557: gap of unknown length
* 125558 148482: contig of 22925 bp in length
* 148483 148582: gap of unknown length
* 148583 180953: contig of 32371 bp in length
* 180954 181053: gap of unknown length
* 181054 181621: contig of 568 bp in length
* 181622 181721: gap of unknown length

```





QY 6172 TTGTAAATTTACATGTAATATGCAATATTTTCCAGCTTTTATATATATAGGCTATGAGCC 6231  
|||||  
Db 115887 TTGTAAATTTTCCATGTAATATGCAATTA-TTGCACATTTTACCTATAGGCTATGAGCC 115829  
QY 6232 CATGTCATATAGAAAGACAGAAATCTAGCTTACACAAAGTTGCACAAATGTATCTAA 6291  
|||||  
Db 115828 CATGTCATATAGAAAGACAAACCTCTAGCTTACACAAATGTGCACAGATGTATCTAA 115769  
QY 6292 GCATTAAGTAATGTGAAACATAGCACTGTAATCTCAAGTCCCTGCTGATGATCAAGT 6351  
|||||  
Db 115768 GCATTAAGTAATGTGAAACATAGCACTGTAATCTCAAGTCCCTGCTGATGATCAAGT 115709  
QY 6352 CAGAATGCAATTAATCTGCTGATTCCTCATCTTTTGTACTACTGTTAGCTGATATG 6411  
|||||  
Db 115708 CAGAATGCAATTAATCTGCTGATTCCTCATCTTTTGTACTACTGTTAGCTGATATG 115649  
QY 6412 CTTTAAAGAAACATTTGGTGGAGCTGTATCCTTTTGTATTTTAAATACATTAATTGA 6471  
|||||  
Db 115648 CTTTAAAGAAACATTTGGTGGAGCTGTATCCTTTTGTATTTTAAATACATTAATTGA 115589  
QY 6472 CATATGCTATATTTTGTGTAAGATGTAGAAATGTAATGTAATGCTTATGCTTATCATC 6531  
|||||  
Db 115588 CATATGCTATATTTTGTGTAAGATGTAGAAATGTAATGTAATGCTTATGATC 115529  
QY 6532 CAGTTGTACAGCTGCAAAATTAATATATATACAT 6568  
|||||  
Db 115528 CAGTTGTACAGCTGCAAAATTAATATATATACAT 115492

RESULT 12  
AK074318 3450 bp mRNA linear pri 15-FEB-2002  
LOCUS AK074318  
DEFINITION Homo sapiens CDNA FLJ23738 fls, clone HEP15081, highly similar to  
ACCESSION AK074318  
VERSION AK074318.1 GI:18676888  
KEYWORDS oligo capping; fls (full insert sequence).  
SOURCE Homo sapiens hepatoma cell\_line:Hepeg2 CDNA to mRNA, clone\_lib:HEP  
clone:HEP15081.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1  
AUTHORS Kawabata,A., Hkiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,  
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,  
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3450)  
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T.,  
Shibahara,T., Tanaka,T. and Nakamura,Y.  
TITLE Direct Submision  
JOURNAL Submitted (14-FEB-2002) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure, Human Genome  
Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan  
(E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
Fax:81-3-5449-5416)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology; cDNA library construction,  
5'- & 3'-end one pass sequencing: Depart of Virology and Human  
Genome Center, Institute of Medical Science, University of Tokyo  
(partly supported by Science and Technology Agency).  
FEATURES  
source  
1. 3450  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HEP15081"  
/cell\_line="Hepeg2"  
/cell\_type="hepatoma"  
/clone\_lib="HEP"  
/note="Cloning vector: PME18SFL3"  
CDs 1. 3342

/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB85049.1"  
/db\_xref="GI:18676889"  
/translation="MSVREELCSVWIEFEVDAQILLEDQELDSKVIYLLNGVETIS  
HPDGKVENLEFGNSFGIIPPTDQKQYMHGIVRRKVDQCFVCAQDDVKRLIHHVKNT  
HKVEEGEIVYVHEHRELDKSGTGHIVIKPRLPRLIMHLEHSIADPTIIEDFL  
TYRFLPSLDVGIKLEMPKIDSLDRKQTRIVLLMNNHNFEPEDGAMTFEVE  
KLEDETKMNGHRLININACIORLMEFGAARGPVPVPSQFSLNKGSEKGFVEGVE  
PCKAADSGLKRGDDIMEYNGNENFNTFMKAVELRNHNTALVKNKINPEKELL  
RTEQKSGVPHIPKIAEKSNNHSHIQHPGDEQSQEGSKKVAANYVSGRANKIRK  
ILDKTRFSLIPKFLPSDGLSQSDSDSVGRHCHSLAIMPIDGLSSSPDLQPT  
TSMDFSNPSDIPDVIVRYFKYDQSCYIISKDTAKVEVFAHVEHGLGASDTYS  
LCEVSTPEGVKORRLLPOFSKLADRIOLNRYILKNNMETLCSDEDAQELVES  
OLSMOLSTIEVATOLSMRDEDFLENIETEVILDLFKNSKGTNHLKREDFVNOE  
TPVVASSEITENQIKRKKIIFKIKALHCECKNFNMFATISGLMLASVARIJGT  
WEKLPSEKYEHLDDIDLEPSRNNAKRNILSSOSMRPILIPLVKKDDITLHEG  
NDSKVDGLNVEFLKRMISKEIQVVRNNSANDPAMMRQCKRRRSLSGSSTNS  
NNLDVQGGHKKRRARRSSILNAKLYEDAOMARKYKQYLSLDVETDEKRCQMSLQW  
EPAYGLTKNLSEKRSKSSSEMPVMSRAGQRTKALHQPVRVQVLOVPVNLHP  
RKSGOTKDPALNTSLPOKYLGTTEESGKKHEDTISVASSLHSPASPOGSPKVG  
STIISPHSKISGOSCPGIGGATLQKRIQDITSTAKRDSPEKATEENRSCENT  
RKRWTSPEFRLEHMLSRERLVNSQEDIDHQAATESCKKVDVGSNIXDEGSAIFN  
SMSQGSNLTNCFYTRFSKRRKTL"  
BASE COUNT 1173 a 620 c 763 g 894 t  
ORIGIN  
Query Match 20.4%; Score 1341.4; DB 9; Length 3450;  
Best Local Similarity 70.0%; Pred. No. 3e-308;  
Matches 1894; Conservative 0; Mismatches 796; Indels 15; Gaps 6;  
QY 483 ATGTCAGTGGAGCGGAGAACTCTGCTGTGAGTGGTGCACAGTGTGAAAGAGCAGG 542  
|||||  
Db 1 ATGTCGTGTAAGGAGAGACACTCTGCTGATGATTTTGAAGTGTGAAAGAGCGTGGA 60  
QY 543 ACCATAGTGTTAAATGATGATGTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 602  
|||||  
Db 61 GCTATTATTCTGTAAGATGGGCAAGAGCTTGACTCATGATGATTTATTTAAACGGCAGT 120  
QY 603 GTGGAAGTGACTTATCCAGATGGAAGAAAGCAAAATCTGTCGATGGGAATGATTTGGT 662  
|||||  
Db 121 GTGGAATTCAGTATCCAGATGGAAGAAATTTGTTTAAAGGAAATGATTTTGA 180  
QY 663 GTCTCTCCCTACATGGAACAAGATACATGAAGAGTGTGTAAGCAAAAGTGTGATGAC 722  
|||||  
Db 181 ATTACTGCCCTCTGATATAGCAGTACATGATTCATGAGCTAAGATGAT 240  
QY 723 TGCAGTTTGTGTCATAGCCGACGAAGATTTACTGCCGATTTCTCAATCAAGTAGAAAG 782  
|||||  
Db 241 TGTCAAGTTTGTGTCATAGCCGACGAAGATTTAGGAAATTTTAAACCATGTGGAAAA 300  
QY 783 AACATGGAAGATTTGAAGAGGAGAGATGTTATGTTGTAAGAACCCGAGAACTT 842  
|||||  
Db 301 AATTAACCATTAAGTGTGGAAGAGGAGAAATGTTATGTTGATACATGCAATCGGAACTA 360  
QY 843 GATCGAATGGAACAAGAAAGGACACATTTGCATCAAGGATCCTCAGAAAGTGTAAAC 902  
|||||  
Db 361 GACCGAGTGGAAACGAAAGGACACATTTGATCAAGGCAACACCTGAGCTGCATTA 420  
QY 903 ATGCATTTGGTGAAGAGCATTTAGTAGTAGATCAATTCATAGAAAGCTTTGCTTG 962  
|||||  
Db 421 ATGCATTTAATAGAAAGCAATTCATCCTGTCGATCAACATTTATTAAGATTTTCTATTA 480  
QY 963 ACCATAGGACTTTCTTTTACCCCAAGGAGTGGGCAAAAGTATTTGAGTGGTTT 1022  
|||||  
Db 481 ACTTACGAGACATTTCTTGAAGTCCCTTGTGATGTTGGGATCAAACTATTTGGATGGTTT 540  
QY 1023 AATGACCCGAGCCTCAGGAGTAAGTGTACACGGGTAGTATTTGTGGTAATAATTCAC 1082  
|||||  
Db 541 AAGATGACAGCTTAAAGATTAAGTGAACGAGATTTGATTTAATGGTAATAATTCAT 600  
QY 1083 TTCAATGACTTTTGAAGAGATCTGCAATGACTTTTGAAGAAATTTGAAACAAAT 1142  
|||||



Db 601 TTTAATGATTTGAAGTGACCTCGCTATGACTCATTTCTAGAGAAATTTGAAAAAT 660  
OY 1143 CTGGAAAGAGAAAAATGGGTGGACCTTAAGCGCTGTGAATATCGCGTGTCTGA 1202  
Db 661 CTGGAGAGATACAAAGATGATGTCTATCCGTTATTTGAATATGCGTGT -CTGCAAG 719  
OY 1203 GCAAAAAAGAGATGATGCTTAACAAACCATCCGAGAAAGCTCTTGGCTTTATC 1262  
Db 720 GCTAATGGAGACAGGTTGTGCT- -GCAAAAGCTTCCCGAGTCCCTCACAATTCAG 777  
OY 1263 TTACTTGAAGGCTCTGAGAAGGATTTGAATCTTTGTGACAGTGTAGATTCAGTAC 1322  
Db 778 CTTAATGAGGAGGTGAGAAGGATTTGGTATTTTGTGAAGAGTAAACCTGTAGAC 837  
OY 1323 AAAGCACTGAAGCAGGCTTGAACCGGGGATCAGATTTAGAACTTAATGCGCAAAAC 1382  
Db 838 AAAGCGCTGATTCAGAGCTGAACAGTGTGATCAGATTTAGAACTTAATGCAAAAC 897  
OY 1383 TTTGAATCATTCAGCTGTCAAAAGCTATGAAATCTTAGAAATTAACACATTTATCT 1442  
Db 898 TTTGAGAAATTTTACATTTATTTGAAGCCGTTGAATTTTGAGCAATATATCTCATTTGCA 957  
OY 1443 ATCACTGTGAACCAATTTATTTATTTAAAGAACTTCTAACAAGATTTGTCAGAAG 1502  
Db 958 CTCACCTGTGAAGCAACATTTTGTGTCAAAAGATTAATTTTGTAG- -GACTGAACAA 1014  
OY 1503 AAAAGAAATGGTCCCCCACCCTCTAAATTTGGTGACATTTAAAAAGCCAGTCTAC 1562  
Db 1015 GAAATATGCTGCTCTCTATATTTCCCAAAATTTGCTGA- -AAAAAAAGTAACTCCCAT 1071  
OY 1563 TCCATTCAGATCTTCTGTAGATGAGAAGAGTATGAGACTTGAAGAAAGTGAACAA 1622  
Db 1072 TCTATCAGACATGTGGCAGAGATATTTGAACAGGCATCAGAGAAAGAAATTAAGAAA 1131  
OY 1623 AAAAGTAAAGCAACACTGTGGAGAGAGACAGCTGAAAAAGTACTCGACAAGACT 1682  
Db 1132 GTTAAAGCAAAATCTGTTTACGTGAGAGAAACAAATCAGAAAGTTTGTGATTAACA 1191  
OY 1683 CGGATAGATATCTGGCAGAGAAACATATCATATTTGGATTTGGTGTGATCTCAAGAT 1742  
Db 1192 CGATTTAGATCTTGGCTCCCAAGATTTAGATGAGGCGCTAAGCAATACAGAT 1251  
OY 1743 GACAGCATAGTAGATTTAAGCGACAGACAAAGCATCCCACTGATTTGCCCTGACGTGA 1802  
Db 1252 GACAGCATGTGGGAACAAGGCATGTAGGCATATGTCTGCTATATGCCCATCCCTGGA 1311  
OY 1803 ACCTTATCATCCAGTATCTGATTTATTTAGCAGTCAATCATGCAATTTTGAATTTCACT 1862  
Db 1312 ACATCTCATCCAGAGCCCTGATCTCTGACAGCTACACAGATATGTTGATTTTTC 1371  
OY 1863 GCTACTCCGACTTGGCAGATCAAGTCTTAAGGTTTAAAGCTGATCAGCAAGCCG 1922  
Db 1372 AATCCCTTCAAGATATCCCTGATCAAGTTATTAAGTTTCAAAAGTGAATCAGCAAAATTCG 1431  
OY 1923 TACATCATGATCTAGAGCACTATACAGCAAAAGAGTGTCTATAGGCTATCAGAGAG 1982  
Db 1432 TACATATATCATGATTAAGACACACACAGCTAAAGAAAGTATTTTATGTCTGTTCATGAA 1491  
OY 1983 TTTGCTTACTGCAACCCCGATCAATATTCATATGTGAGGCTCTGTCAACCTGAG 2042  
Db 1492 TTTGGTTTGAACCGGTGATCCGACATATTTCTCTGTGAAGTTTCTGTACTCTCCGAG 1551  
OY 2043 GAGATATCAACAAAGAGACTTCCAGATCAGCTTTCCAACTTGGCAGACAAATACAA 2102  
Db 1552 GAGTATATAAAGAGAAACCTTCAGATCAATCTCCAAATTAAGTAAATTCAA 1611  
OY 2103 CTGAGTGAAGTATTTATCTGAATAACATGAAACAGAAATCTTTTGTAGATGAA 2162  
Db 1612 CTCAATGGAAGGATTTACTTTAAAAATTAACATGAAACAGAAATCTTATGTAGATGAA 1671  
OY 2163 GATGCTCAGAGTGTGTGAGAGAGTCAAAATTTCCCTCCCTTACGCTCAGCACTGTGAA 2222  
Db 1672 GATGCTCAAGAACTAGTTAAGAAAGCAGCTATTCATGCTGACGCTCAGTACCATTTGAG 1731

OY 2223 GTTGAACACAGCTCTCTATGGAATTTTGAACCTCTTTCGACATTTGAACTACTGAA 2282  
Db 1732 GTGGCCACCAGCTGTCAATGAGGAGCTTTGATTTTGTGTAATATTTGACCGACTGAG 1791  
OY 2283 TATATGATGATTTATTTAACTCAGATCAAAACAGCTGTGCCAAGTGAAGATTT 2342  
Db 1792 TACATGATGACCTTTTAAAGTTAAATTCGAAACAGAAATTCATTTGAAGAGCTT 1851  
OY 2343 GAAGAATCATTAACCAAGAAATTTTGGTATGATCTGAATTTCTCAGAGAAACAA 2402  
Db 1852 GAGGACATTTGTAACCAAGACATTTCTGGTGTCTCAGAAATTTTAACTGAAGCAAT 1911  
OY 2403 CAGCTGAAGAGTGAAGATCAATTTATGATCAAGATTAACACTGACCTGATAGGAA 2462  
Db 1912 CAGCTGAAGAGATGATTAAGATTTATTAATTTGAAATTTGACCTGATGTCAGAA 1971  
OY 2463 TGCAGAAATTTTAACTCAATGTTTGAATCATCAGTGGCTTAACCTGGCAGCAGTGA 2522  
Db 1972 TGTAAAGATCTTCAATTCATGTTTGCATATATATAGTGCTTGAACCTGCTATGACA 2031  
OY 2523 AGACTGGAACGACCTGGAGAAACTTCCCAATTAATAGCAAAACATATTTCAAGATCTC 2582  
Db 2032 AGACTGAGAGAACTTGGGAAAGTTACCAAGCAAAATAGCAAAACATCTTCAAGATCTA 2091  
OY 2583 CAAGACTGTTTATCTTCCAGAAACATGGAATATGCTAATGTTCTCATATAGTCAA 2642  
Db 2092 CAAGACATTTTATCATCTCTAGAAACATGGAATATGAAATTTCTTATAGTACAA 2151  
OY 2643 AATCTCAACCTCCCAATCCCTATTCCTATTCAGATTATCAAAAGATCTCACCTCTCT 2702  
Db 2152 AGTATCGGCTCCATTTATTCACCTTCCCTGTTGTCAAGAAATATGACATTTCTA 2211  
OY 2703 CACGAAGAAATGCTCAAAAGTATGACGGGCTGCTCAATTTTGAAGACTAAGATGAT 2762  
Db 2212 CATGAAGAAATGCTCAAAAGTATGATGTTATGAATTTTGAAGATTAAGAAATGAT 2271  
OY 2763 GCAAAAGAAATGCTCAGCTGGCAGATGGCTTCAAGTGAACATGAGCCCTGCCATG 2822  
Db 2272 TCCAGAAATTCGCAAGTGTTCGAAATGCTGTGCTAATGAGCCAGCTATGATG 2331  
OY 2823 TTTGAGACTGGAAGAAATGCGGAGTTTGGGCTCTCTCAGCCAGGTTAGTACAAT 2882  
Db 2332 TTTGACAGAGGGAAGAAAGAGTGTGCGGAGTCTGGGTCACTGATGACGAACACAAAT 2391  
OY 2883 GCAAGCTGATGATGTTGCTCAGACAGTGTGCTATTAAGAAAGCGGATGCTGATTC 2942  
Db 2392 TCAACATGCTGATGT- -TCAGGAGGTGCTCACAAAAAAGGGCAGCGCAGCTCT 2448  
OY 2943 TTTCTCAATGCCAAAAAGCTTTTGAAGATGCCCAATGCTGAAAAGTGAAGCAGTAC 3002  
Db 2449 CTGCTTAATGCCAAGAGCTTATGAGGATGCCCAATGCAAGGAGTGAAGCAGTAT 2508  
OY 3003 CTTTCAATTTGAGAGTGAATGAGCAGAGAGAGTCTTGAAGATTTATCTGACAGTGT 3062  
Db 2509 CTTTCCAGTCTCATGTATGAGACAGATGAGGAGAAAGTTCCAGATGATGTATTCAGTGG 2568  
OY 3063 GAGCAGCAACAACACATTTGCTTAAGAAATCTGTGACAAAGCCGTGTCAA- -ATCC 3119  
Db 2569 GAGCTGACATATGATGATGCTGACCAAGAAATTTAAGTGAAGAAAGATCAGCAATATCT 2628  
OY 3120 GAGACCTTCCATAGCTCCAAAGGAGGTCACAAACAGAAAGCTAGTCCCTGCCAG 3179  
Db 2629 GAAATGCTTCCATGCTATGAGGTGAGTGGCCCAAAACATTAAGGCCCTTGATCAA 2688  
OY 3180 CCCC 3184  
Db 2689 CCCC 2693

RESULT 13  
AC118835/c  
LOCUS AC118835 185108 bp DNA linear HTG 18-JUL-2002

DEFINITION Rattus norvegicus clone CH230-262L23, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 63 unordered pieces.  
ACCESSION AC118835  
VERSION AC118835.4 GI:21747243  
KEYWORDS HTG: HTGS.PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 185108)  
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alshrocks,S.L., Amaralunga,H.C., Are,J.R., Ayala,M., Banks,T.,  
Barberia,J., Benton,J., Bimge,K., Blankenburg,K., Bonini,D.,  
Bouck,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Bui,X., Burkett,C., Butrell,K.L., Byrd,N.C.,  
Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabriel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogues,M., Hollaway,C., Hollins,B.,  
Homsli,F., Howard,S., Huber,J., Huik,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kretovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.,  
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,  
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Meheswaril,M., Mapa,R., Martin,R., Matlindale,A., Martinez,E.,  
Messer,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Moser,N., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokwenko,S., Oguh,M., Okwuon,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peeters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Saverly,G.,  
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,  
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Swatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Usmani,K., Vasquez,L., Vera,R., Villalob,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,  
Williams,G., Williamson,A., Wlaczek,R., Woodson,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 185108)  
AUTHORS Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (21-APR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 185108)  
REFERENCE Direct Submission  
AUTHORS Worley,K.C.  
TITLE Submitted (18-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 14, 2002 this sequence version replaced gi:20330712.  
COMMENT ----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GVTY

Center clone name: CH230-262L23  
----- Summary Statistics  
Sequencing vector: Plasmid:  
Chemistry: Dye-terminator Big Dye: 100% of reads  
Assembly program: Phrap: version 0.990329  
Consensus quality: 127499 bases at least Q40  
Consensus quality: 136160 bases at least Q30  
Consensus quality: 141175 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a "working draft" sequence. It currently  
\* consists of 63 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1  
\* 1022: contig of 1022 bp in length  
\* 1023: gap of unknown length  
\* 1123: contig of 1015 bp in length  
\* 2138: gap of unknown length  
\* 2237: gap of unknown length  
\* 2238: contig of 186 bp in length  
\* 3424: gap of unknown length  
\* 3524: contig of 1257 bp in length  
\* 4781: gap of unknown length  
\* 4881: contig of 1190 bp in length  
\* 6071: gap of unknown length  
\* 6171: gap of 1036 bp in length  
\* 7207: gap of unknown length  
\* 7306: contig of 1333 bp in length  
\* 7307: gap of unknown length  
\* 8640: gap of unknown length  
\* 8640: contig of 1500 bp in length  
\* 8740: gap of unknown length  
\* 10240: contig of 1257 bp in length  
\* 10339: gap of unknown length  
\* 11586: contig of 1864 bp in length  
\* 11597: gap of unknown length  
\* 11696: contig of 1864 bp in length  
\* 11697: gap of unknown length  
\* 13561: gap of 1297 bp in length  
\* 13660: gap of unknown length  
\* 14957: gap of unknown length  
\* 15057: gap of 1189 bp in length  
\* 15058: contig of 1189 bp in length  
\* 16246: gap of unknown length  
\* 16247: gap of 1827 bp in length  
\* 16347: contig of 1827 bp in length  
\* 18174: gap of unknown length  
\* 18273: gap of 1500 bp in length  
\* 19773: contig of 1064 bp in length  
\* 19774: gap of unknown length  
\* 19873: gap of 1064 bp in length  
\* 20937: gap of unknown length  
\* 20938: gap of 1657 bp in length  
\* 21037: gap of unknown length  
\* 22694: gap of unknown length  
\* 22695: gap of 1589 bp in length  
\* 22795: gap of unknown length  
\* 24383: gap of 1221 bp in length  
\* 24384: gap of unknown length  
\* 24484: contig of 1221 bp in length  
\* 25704: gap of unknown length  
\* 25705: gap of 2353 bp in length  
\* 25804: gap of unknown length  
\* 28163: gap of unknown length  
\* 28164: gap of 1138 bp in length  
\* 28264: gap of unknown length  
\* 29401: gap of 2133 bp in length  
\* 29502: gap of unknown length  
\* 29502: contig of 2133 bp in length  
\* 31634: gap of unknown length  
\* 31733: gap of 2188 bp in length  
\* 31734: gap of unknown length  
\* 33921: gap of unknown length  
\* 33922: gap of 1736 bp in length  
\* 34021: gap of unknown length  
\* 35757: gap of 1657 bp in length  
\* 35758: gap of unknown length  
\* 35857: gap of unknown length  
\* 37514: gap of 2116 bp in length  
\* 37515: gap of unknown length  
\* 37615: gap of 2116 bp in length  
\* 39730: gap of unknown length  
\* 39731: gap of unknown length  
\* 39830: gap of 1909 bp in length  
\* 39831: gap of unknown length  
\* 41739: gap of 2155 bp in length  
\* 41839: gap of unknown length  
\* 43994: contig of 2155 bp in length  
\* 44094: gap of unknown length

44095	44581:	contig of 1487 bp in length
44582	45681:	gap of unknown length
45682	47275:	contig of 154 bp in length
47276	47375:	gap of unknown length
47376	50938:	contig of 3563 bp in length
50939	51038:	gap of unknown length
51039	53405:	contig of 2367 bp in length
53406	53505:	gap of unknown length
53506	55385:	contig of 1880 bp in length
55386	55485:	gap of unknown length
55486	57173:	contig of 1688 bp in length
57174	57273:	gap of unknown length
57274	58404:	contig of 1367 bp in length
58404	58940:	gap of unknown length
58941	61253:	contig of 2313 bp in length
61254	61353:	gap of unknown length
61354	63243:	contig of 1890 bp in length
63244	65343:	gap of unknown length
65344	65848:	contig of 2505 bp in length
65849	65948:	gap of unknown length
65949	70327:	contig of 4319 bp in length
70328	70427:	gap of unknown length
70428	72825:	contig of 2388 bp in length
72826	72926:	gap of unknown length
72926	74661:	contig of 2036 bp in length
74662	75061:	gap of unknown length
75062	78369:	contig of 3308 bp in length
78370	78469:	gap of unknown length
78470	81136:	contig of 2667 bp in length
81137	81236:	gap of unknown length
81237	84350:	contig of 3114 bp in length
84351	84500:	gap of unknown length
84501	87806:	contig of 3356 bp in length
87807	87906:	gap of unknown length
87907	91476:	contig of 3570 bp in length
91477	91576:	gap of unknown length
91577	94491:	contig of 2915 bp in length
94492	94591:	gap of unknown length
94592	99207:	contig of 4616 bp in length
99208	99307:	gap of unknown length
99308	103122:	contig of 3815 bp in length
103123	103222:	gap of unknown length
103223	107224:	contig of 4002 bp in length
107225	107324:	gap of unknown length
107325	110425:	contig of 3101 bp in length
110426	110525:	gap of unknown length
110526	114415:	contig of 3890 bp in length
114416	114515:	gap of unknown length
114516	117090:	contig of 2575 bp in length
117091	117190:	gap of unknown length
117191	123776:	contig of 6586 bp in length
123777	123876:	gap of unknown length
123877	129578:	contig of 5702 bp in length

QY	4755	ATTGCCCTGGCAGCTTTTCAGACTTTGTTGCTTGAATATGCACAGTGCAGCAATCTTCGACC	4814
Db	148665	ACTGCCCTGGCAGCTTTTGAGAC-TTGTGTGCTGAAAGTCACACCAGCAGAGCTCTGACG	148607
QY	4815	TCGCCATGTCTGCGCTCCGCACATCACACAGTATCATTCCAAAATCCAGATCATCAAA	4874
Db	148606	-----TCTGCTCTTCCACGTCGACGACGATCATTCCTCAAAATTCGMAATGTC--A	148559
QY	4875	CAGATGATTCACCTCTGCGCTCCACCTCTCAATGCGCTGGAAGGATTTTATTAACTTCCT	4934
Db	148558	CAGGCGACCTCCCTCTGGCGGACACTCTCCATGCGCTGGAAGGA--GTTCCAGCTTCCCT	148501
QY	4935	TTTAAATTTCAATCCAGTCCCTAGCAGCTGTGATCTCATTTGGGATATAGAAAACCTTGCCA	4994
Db	148500	CTTTAAATTTTCAGTCCAGTCCCTAGCAGCTGTGATCTCATTTGGGAAAGAGAAAACCTTGACCT	148441
QY	4995	TTGAATCTACTTGGGGCCCTTTAAACCACCAAGAGACAAAAGAAAACATGAATCTCTT	5054
Db	148440	TTGAATCTACTTGGGGGCTCTCGAATCCACCAAGAGACAAAAGAGACAGATTA--GAATCTCTTA	148382
QY	5055	CAGTCACTGCTGTGTCCACTTGTTTACAATGTCTCTTTTAAAAAAAATATGAGTTTA	5114
Db	148381	GGATCAAGTGTCTTGCCCACTTGTTTACAATATCTTT-----AAAAAAATAGCTTTTA	148330
QY	5115	AAGATTTTGTTCAGAGATTAAT-ATPATCATTTTAATGATTTACGATATATTTTAAAC	5173
Db	148332	AAGATTTGTGTTCAGAGATTAATGTATATTCATTTAATATTAACGATATATTTTGAAC	148270
QY	5174	CTTAAATAGGGTTGGCAGCCCTG-CTTCTCGAAAAACCAAAATATGCGGACAGGAGTGTGC	5232
Db	148269	CTTAAATAGGGATTTGGCAGCCCTGAGATTTCTGAAAAACCAAAATATGCGGACAGGAGTGTGC	148210
QY	5233	CACACCAAGAAGAGGGAAGACGCTGCGCTGTGACCCGCTGCTCCCATGCTCTCTGCTCT	5292
Db	148209	CGCACCAA-----GGGAAGACCTGGGACGAGACCTGTCTCCCATGCTCTCTGCGCT	148156
QY	5293	CACCCGCAAGTGCCTATCTCTGGAAGATTAATGTTAGCCAAAT-----AATACCAAG	5347
Db	148155	TCCCGTGAAGTGCCTATCTCCAGAGATCAAAAGTTAGCCAAATCTTAATATCCAG	148096
QY	5348	ACACTCATCTGCTCTTCCCACTGATGATGGGTTCTTCTGTAAACTGTTTGCACATGG	5407
Db	148095	ACTCTCATCTGCTCTTCCCACTGATGATGGGCGGCTCTTAATGTAAACTGTTTGCACATGG	148036
QY	5408	CCAGGGAAGGCACTAGSAGCCCTGTGTGTCGCTGAGCCCTATAGGAGGAGGACGGTGT	5467
Db	148035	CCA-GGTGGGAACATAGSAGCTGTGTGCTGTGAGCCCTAT-GAGGAGGAGATGGCT	147978
QY	5468	CATTGGCGATGTGCTGCTTCATAGATGATGATGCAAAACCCCATTTTTTAAAGTTA---	5524
Db	147977	TATTGAGAAGTGTGCTGCTTCATATCAAAATGATGCAAAACCCCTGTTTAAAGTTATGG	147918
QY	5525	-----TATTCTTGATTTTTGTAAATTAAGAGGTG	5556
Db	147917	GTTTTTTTTTTTGTGTTTGTGTTTGTGTTTTGTTTTTTTTTTTTTTTCAAGTTTGTTAATTT	147858
QY	5557	AGTTTTCTTTTTGTTTTTTGTTTTTTTTTAAAGAGAACAATTTATTAACGTAGATACATT	5616
Db	147857	AGGTTTTTATAGACTTTTTGTTTTTATTTTTTAAAGAGAACAATTTATTAACGTAGATACATT	147798
QY	5617	GCAGTGAAGCAGCTTGGGATCTTGGAGCTTAATGCCAGCTGTTTATATCTGCTCTTCAAG	5676
Db	147797	GCAGTGT-AAGCAGCTTGGGGGTTGGGAAAAATATCCAGCTGTTTGTATCTCTCTTCAAG	147739
QY	5677	ACAGGCTCCCTTTATGTAATTTGCAATTAAGGAAATTAACAAGCCCTTTAAAGGTATTAAG	5736
Db	147738	ACAGGCTCCCTCTGTGTGATTGGACACAGAG-----TGTCTTTAAACATGTTTAAG	147687
QY	5737	ATCAAAAACCTGTGTTAGACATGCCAGCCCTTTCAGAGCCAGGTTATGTCACCAAACTAAC	5796
Db	147686	GTCAAAAACCTGTGTTAAAGTATACCAAGCCCTTTAAAGGTGGTACGTA-CCGAAGCTGAC	147628





AUTHORS  
TITLE  
JOURNAL

## COMMENT

Worley, K.C.  
Direct Submission  
Submitted (09-FEB-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Oct 14, 2001 this sequence version replaced gi:11079356.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: MAM

Center clone name: RP23-361K18

Summary Statistics

Sequencing vector: M13; L08621

Chemistry: Dye-primer Body: 68% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 139795 bases at least Q40

Consensus quality: 171357 bases at least Q20

Estimated insert size: 165623; sum-of-coverage estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 2.4x in Q20 bases; sum-of-coverage estimation

NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

NOTE: This is a "working draft" sequence. It currently

consists of 36 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 10583: contig of 10583 bp in length

10584 10683: gap of unknown length

10684 20426: contig of 9743 bp in length

20427 20526: gap of unknown length

20527 29062: contig of 8536 bp in length

29063 29162: gap of unknown length

29163 36210: contig of 7048 bp in length

36211 36310: gap of unknown length

36311 42380: contig of 6070 bp in length

42381 42480: gap of unknown length

42481 47818: contig of 5338 bp in length

47819 47918: gap of unknown length

47919 53861: contig of 5943 bp in length

53862 58028: gap of unknown length

58029 58128: contig of 4067 bp in length

58129 61169: gap of unknown length

61170 61269: gap of unknown length

61270 65935: contig of 4666 bp in length

65936 66036: gap of unknown length

66037 68826: contig of 2791 bp in length

68827 68926: gap of unknown length

68927 71855: contig of 2929 bp in length

71856 71956: gap of unknown length

71957 76368: contig of 4413 bp in length

76369 76468: gap of unknown length

76469 79877: contig of 3409 bp in length

79878 83881: contig of 3904 bp in length

83882 83981: gap of unknown length

96578 96677: gap of unknown length

96678 100656: contig of 3979 bp in length

100657 100756: gap of unknown length

100757 103738: contig of 2962 bp in length

103739 103838: gap of unknown length

103839 106689: contig of 2851 bp in length

106690 106789: gap of unknown length

106790 109708: contig of 2919 bp in length

109709 109808: gap of unknown length

109809 113553: contig of 3745 bp in length

113554 113653: gap of unknown length

113654 116605: contig of 2952 bp in length

116606 116705: gap of unknown length

116706 120077: contig of 3372 bp in length

120078 120177: gap of unknown length

120178 122520: contig of 2343 bp in length

122521 122620: gap of unknown length

122621 125560: contig of 2940 bp in length

125561 125660: gap of unknown length

125661 128466: contig of 2806 bp in length

128467 128567: gap of unknown length

128568 130902: contig of 2336 bp in length

130903 131002: gap of unknown length

131003 134453: contig of 3451 bp in length

134454 134553: gap of unknown length

134554 137254: contig of 2701 bp in length

137255 140707: contig of 3353 bp in length

140708 140807: gap of unknown length

140808 143398: contig of 2501 bp in length

143399 143498: gap of unknown length

143499 146274: contig of 2776 bp in length

146275 146375: gap of unknown length

146376 148466: contig of 2092 bp in length.

Location/Qualifiers

1.148466

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="RP23-361K18"

BASE COUNT 39949 a 31869 c 32369 g 40737 t 3542 others

ORIGIN

Query Match 12.9%; Score 845.6; DB 2; Length 148466;

Best local similarity 75.4%; Pred. No. 3.5e-190;

Matches 1499; Conservative 0; Mismatches 360; Indels 128; Gaps 31;

QY 4536 GAGATTAACAGTTCTGCTGTTGAGCAGACCTTTCTGAGAGCAGCAGC 4595

DB 72227 GAGACGACAGAGTGTGCTATTGAGCAGAGCTCTTTTCCAGATGACACCC 72286

QY 4596 CTGAAGAGAGCAGCAGAGAGAGCTCTGAGCATTTGAGAGCTTCTGAG 4655

DB 72287 CAAAGGACAGACAGCAGAGAGATGTCACAGCAGCTTGAGCAGCAGC 72346

QY 4656 GAGCGTGACAGTTTCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4712

DB 72347 GATGCTGACAGAGTTTCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 72406

QY 4713 CCCCTTCT 4772

DB 72407 CCCCTTCT 72465

QY 4773 AGACTTGTGCTGTAATGACAGATGACAGCAATCTTGAAGTCCAGCTTCTCTCT 4832

DB 72466 TGCG-CTGTGTTAGAAATACAG-----CCGAGAGCTCTCTCTCTCT 72506

QY 4833 CCACATACACAGATATTCCTCAATTCACATCATCAGACAGATGATCTCTG 4892

DB 72507 CCACGTTTCCGATATTCCTCAATTCACAGATGATGATGATGATGATG 72563

QY 4893 CTGACATTCATACGCTGAGAGATTTTAAATCTCTCTCTCTCTCTCTCTCTCT 4952

DB 72564 GGACATCTTCATGCTCTGAGAG--GTTTCAATCTCTCTCTCTCTCTCTCT 72621

OY	4953	CCACACTTATATCTGATTTGGATATATAGAAAAACCTGCCCTTGAACTACTTGGGGCT	5012
Db	72622	CCATACCTCGATCTCATTTTGGAGATATAGAAAACCTTGGAATCTACAGG----	72676
OY	5013	TTAACCCACAAGGAAGCAAAAGAAAACATGAATCTCTTGAGTACAGTCTGTGCA	5072
Db	72677	-----AAGACACAGACATATATGAATCTTAGAGTA-TGGCTTGTGCA	72719
OY	5073	CTTGTTTACATGTCTCTCTTTTAAAAAAAATGAGTTTAAAGATTTTGTTCAGAG	5132
Db	72720	CGTGTTTCAACATCTCT-----TTAAATTAACCTTAAAGATGTGTTCAGAGG	72769
OY	5133	TAAAT-ATATATCATCTTAATGATTAAGT-----ATTATTTTAACTTAAAGTGGCT	5186
Db	72770	TAAAGTTACATCTATTTAGTATATATTTATATATCATTTTGAACCTTTAAGTAGATT	72829
OY	5187	GCCAGCTGG-GTTCCTGAAAAACCAATATGCGGACAGGGGTGGCCACACCAAGAA	5245
Db	72830	GCCACCTGTGTCTTGTGAAAAACCAATATGCGAAGAAAGGHTTGTGTGGCCAA--	72887
OY	5246	CGGGAAGACCTGGCTGTGGACCTTGGCTTCCATGTCTTGTGCTACCCGGAATG	5305
Db	72888	--GAAGACCTGGGAGATGTCTGTCTTCCATGCTCTTGGCTTGGCCAGTAAATG	72944
OY	5306	CCCTATCTCGAAGATATAAATGTAGGCAATTA-----ATACAGACACTCATCTGC	5360
Db	72945	CCCTATCCAGACATATCAATATGTACCANTACTTATGATACCAAGACTTCTACCTGC	73004
OY	5361	TCTTTCCCACTGGATGGGGTTCTCTGTAAACTGTTTGCACATGCGCAGGGAGGAA	5420
Db	73005	TCTTTCCCACTGGATGGGGTTCTCTGTAAACTGTTTGCACATGCGCAGGGAGGAA	73060
OY	5421	CTAGACACCTGTGTGCTGCTGACACCTTATGAGGACGAGAGGCTATTGGCCAGAT	5480
Db	73061	CTTGACCTGTGTGTGCTGCTGCTGACACCTGTG- GAGGACGAGATGGTGTGTTGAACCT	73119
OY	5481	GTCCTGCTTCATGTGATGATGGCAGAAACCCATTTTAAG--TTATATTTCTTTGAT	5537
Db	73120	GTCCTGCTTCATGTGATGATGGCAGAAACCCATTTTAAG--TTATATTTCTTTGAT	73179
OY	5538	TTTGTATTTAGAGGTGAG-----GTTTGTTTTGTGTTTGTGTTTGT	5585
Db	73180	TTTGTATTTAGAGGTGAG-----GTTTGTTTTGTGTTTGTGTTTGT	73239
OY	5586	TTAAGAGAACTTATATACGTGATAGCATGTGCATGAAGACAGCTTGGAGTGTGAGC	5645
Db	73240	TATTTTAAAGAACATATACGTGATAGCATGACAGTG-AAGAGCTTGGGGTGTGGAGT	73298
OY	5646	TATATCCAGCTGT-TATATGCTCTTTTCAGACAGCTCCCTTATTTGAATTTGCAATTA	5704
Db	73299	GATATCCAGCTATTCGTGATTTGTTCTTTCAGACAGCTCCCTTCTGAGATTTGACACA	73358
OY	5705	GGGATTAACAAGCCTTTTAAAGTGATTAAGATTAACAAACCTGGTTTGACATGCGACG	5764
Db	73359	GAGAGT-----ATGTCTTTAAACATAT-AGAGGTACAAACATGGTATCATGCACTAGCC	73413
OY	5765	TTTGCAGGACAGTTAGTCACCAAGACTAACCCTCAAGTGAGCTTTATGAGAGCTGCATA	5824
Db	73414	TTTGCAGGACAGTTAGTTAGTCACCAAGGCTGATGTAGTGCTTTGGAGATGTGCACA	73472
OY	5825	TAGACAAAGGCTTAATGTAGCAACATGTGCTCACAGCTGCTATTAACCTATATGACT	5884
Db	73473	TGGAACAAAACCTTAAGTAGTATGTCTGTTCATAGCTGCGTTAAGCCCTAAGATGACT	73532
OY	5885	G-AAATGACCCCTCACACTTATTTTGTGT- GTTTTCCACAGCACTCGGAAATGTAG	5942
Db	73533	GAAATGACCCCTCGCGGTCTATTTTGTGTCTTTTTCGAGGACCTTCAGAAAATGTAG	73592
OY	5943	GCTGCCAATCTGATAGTACTCAATGTGAGGACCTGCTGCTTGAGATTTTTCAT	6002
Db	73593	GCTGCCAATCTCAAGTAGTACTCAAGTAGTGGAGGACCTGCTGATCTTGGG--TTTTCAT	73649

QY	6003	TAAATTCAGCTATCATATTTGATGCTAGTATTAAGCTAAATAGCTCAATTTTAAAGT	6062
Db	73550	TAAATTCAGCTATCATATTTGATGCTAGTATTAAGT-GCCCTCAATTTTAAAGT	73708
QY	6063	GGAATTCGAGTCTTTTTCACATGATCAACAAAGTCAAGTGCCTTATTTAAATATTCCT	6122
Db	73709	CAAAATTCGAGTCTTTC-TCAGTGCACAAACAAATGCAAGGCTTATTTAAATATTCCTCN	73767
QY	6123	TCTGATATCATGCGATTTGCTCTCTTGCTATATACATTTGCAATTTATGCAATTTGTAATTTT	6182
Db	73768	TCTGGGCCAGGGCAATTTGCTACTCATGATATCACACTGCACTGCTGTTATATATCA	73827
QY	6183	ACATGTAATATGCAAT-----TATTTGCCAGTTTATTTATATAGGCTATGGA	6228
Db	73828	ACATGTAACGTATATAAATATGTAATGTGCATTTTGGCAGTTTATTTCAATAGGCTATGGA	73887
QY	6229	CCTCATGTGCAATATGGAAGAAGACAAATCTAGCTTACCACAAAGTTGCCAATATGTTATC	6288
Db	73888	CCTCATGTGCAATATGGAAGAAGACAAACCTAGCTTACCACAAATTTTACAGAGTGTATTC	73947
QY	6289	TAGAGATTAAAGTAATTTAGAACATAGGACCTGCTATATCTCACTGGGCTGTGATGTCAA	6348
Db	73948	TACACATTATAGTAATTTGTAGAACACAGGACCTCTATCTCAGCTGCGCTGTGATGTCAA	74007
QY	6349	GTCGCAATGTACAAATTAACGTGTAATTTCTCATATCTTTTGATATCTACTTGTACCTGTA	6408
Db	74008	GTCGCAAGTAATGTA-AGTTAACTGCTGATTTCCCTTAATCTTTTATATCACTGTTATCTGTG	74066
QY	6409	TCTCTTTTGAAGAAGCATATGGGAGAGTGTGTAATCCCTTTTGTAAATTTTAAACAAATAT	6468
Db	74067	TATCTTTTGAAGAAGACACTGGTGGAGGCTGT-CCCTTTGTATTTCTACATACAACAAT	74125
QY	6469	GTACATA 6475	
Db	74126	GTACATA 74132	

Search completed: February 18, 2003, 15:52:10  
Job time : 17748 secs

**BLANK PAGE**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2003, 21:17:19 ; Search time 57 seconds

(without alignments)  
2528.167 Million cell updates/sec

Title: US-09-911-826a-2

Perfect score: 7721

Sequence: 1 MKPLAIPANHCWGGQEKHS.....PYSGSGFSTEDEDEQVSAY 1499

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1798.5	23.3	1305	T23314	hypothetical prote
2	540	7.0	1213	S42368	guanine nucleotide
3	455	5.9	96	T17209	hypothetical prote
4	354.5	4.6	1596	A41216	guanine nucleotide
5	325.5	4.2	1095	PC1114	SKCPC25 protein -
6	321.5	4.2	1048	S64758	SCD25 protein (ver
7	310.5	4.0	1250	S14177	SCD25 protein (ver
8	309	4.0	1260	S28407	guanine nucleotide
9	307.5	4.0	1336	S25716	Ras guanine nucle
10	303.5	3.9	1333	A37488	Ras guanine nucle
11	302	3.8	1333	S30356	CD25 protein homo
12	295	3.8	1086	JC7732	C3G protein, long
13	290	3.8	1275	A38985	nucleotide exchang
14	288	3.7	1244	S29083	guanine-nucleotide
15	282.5	3.7	1589	RGBYC5	cell division cont
16	280	3.6	1383	T13052	guanine nucleotide
17	277	3.6	911	S28098	guanine-nucleotide
18	267	3.5	1189	T42726	guanine nucleotide
19	257	3.3	596	T29361	hypothetical prote
20	254	3.3	987	T40241	hypothetical prote
21	245.5	3.2	1297	S25714	son-of-sevenless-2
22	244	3.2	754	T12453	hypothetical prote
23	238.5	3.1	768	A56234	ral guanine nucle
24	236.5	3.1	852	S28415	guanine nucleotide
25	233.5	3.0	638	T21321	hypothetical prote
26	232	3.0	1435	BVBL1	guanine nucleotide
27	229.5	3.0	1745	A46431	flight junction-ass
28	222	2.9	2346	T13829	tpi homolog - frui
29	219.5	2.8	1577	T19722	hypothetical prote

30	218	2.8	1736	2	A47747	tight junction pro
31	217.5	2.8	777	2	T08659	ral guanine nucleo
32	216.5	2.8	2094	2	S33124	tpi protein - huma
33	212.5	2.8	1464	2	T13716	bazooka gene prote
34	211	2.7	1875	2	S38173	myosin-like protei
35	206.5	2.7	1837	2	T41023	probable nuclear p
36	206.5	2.7	2845	2	T49505	adenomatous polyp
37	206.5	2.7	3122	2	T17202	DNA-directed DNA p
38	206	2.7	1538	2	T29095	cardiac muscle fac
39	203	2.6	2253	2	T30336	nuclear/mitotic ap
40	201	2.6	2361	2	T25752	hypothetical prote
41	200.5	2.6	842	2	T16198	hypothetical prote
42	200	2.6	1111	1	A42640	kinesin-related pr
43	197.5	2.6	1023	2	A21843	ATP-dependent dsb
44	197.5	2.6	2101	2	A42184	nuclear mitotic ap
45	197	2.6	1066	1	A48669	kinesin-related pr

## ALIGNMENTS

```
RESULT 1
T23314
hypothetical protein T14G10.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T23314; T24919
R:Wild: A.
submitted to the EMBL Data Library, February 1996
A:Reference number: T23314
A:Accession: T24919
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1305 <W12>
A:Cross-references: EMBL:269664; PIDN:CAA93519.1; GSPDB:GN00022; CESP:T14G10.2
A:Experimental source: clone K04D7
R:Wild: A.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z19954
A:Accession: T24919
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1305 <W12>
A:Cross-references: EMBL:268880; PIDN:CAA93100.1; GSPDB:GN00022; CESP:T14G10.2
A:Experimental source: clone T14G10
A:Gene: CESP:T14G10.2
A:Map position: 4
A:Introns: 450/1; 463/2; 696/2; 763/2; 843/2; 935/3; 1012/1; 1091/1; 1143/1; 1189/2;

Query Match      23.3%; Score 1798.5; DB 2; Length 1305;
Best local Similarity 35.3%; Pred. No. 5e-84.;
Matches 459; Conservative 216; Mismatches 435; Indels 191; Gaps 36;

QY 80 ETVVDSDDDDDEED-----IERASDPLMS-RDVIYRCLERDPIDRDTDDELEHFNOL 133
DB 121 ETVSNEGADSDDEDEGSMPSOESSGGFMRLDSVRECLEREPSERNSDGLAVLDFWQH 180
QY 134 PAFANMTSVYRELCAMVFAVERAGTIVLNDGEELDSVILNGSEVETYPDGKAEIL 193
DB 181 SARFALPMSTIKRQLCKMFAVAVNDAGTVLHNEKIDSYSVINGVEVVRKSGERVEY 240
QY 194 CKNSTGVSFTMDKEYKGMVTKVDDCOFVCAI00DYCRILNOVEKNMOKVEE--EGET 251
DB 241 KLDSDSGAEPTPATQIHIGEMRTMVDDCEFLVLEHRDFCSIMSTIGDHIEKRDGLGVEV 300
QY 252 VAMKEHRELDRTGTRGHTVITKGTSEELTMHLYEE--HSVVDPTFEEDPLTYRTFLSSP 309
DB 301 VSEVERFTV---GTGCGQVLTKGRPKDLHNLVYDERDN--VDHYYVDFFLTTRVFIIDP 356
QY 310 MEVQKLEWFNDPSSLRDKVTRVLLVNNHFNDFEGDPMATFLEFEFNLEBREKMGCH 369
DB 357 TTFIEKMLMFADSIYRDKVARLVLLVNNHFNDFEINDEMMNLEFEBGALERDGHNSQ 416
```

Qy	370	LRNINICAKKARRMLTMTLTKPKSRREAPLPEFTLLGSEKGFJFVDSVDSOGSKATEADLTK	429
Db	417	LSLNLINICSVKAPROVILLR-RKDDKMMRLVAGGSEGSNSYVAENFPDITSAKREYKR	475
Qy	430	GDQJLEVNGCNFENIOLSKAMEILRNHNLSTVTNLFVEKELTRLSEK-----RN	483
Db	476	ADEMLVENOQSAYLAKSKAKKEDDITLSLTMLKNVYLGKETTGGKEHKKPKNGTSR	535
Qy	484	GAPLPIKIGDIKAKSRSTPDLAVDEQYIGLEXYNKKSKANTYGNRRKJLKL-----	537
Db	536	GA-GIPWVIVPHKTS-----ITGKKSSTTSKS---GMEKMLTLLKSKED	578
Qy	538	-----DKTRISLIPKREYNDIGIGOSODDSIYGLRQTKHIFALPEVSGTLLSSNPDLLOS	592
Db	579	SMDFTEAKISSADLRP-----SRSPNDITS-604	604
Qy	593	HHRLILDF--SATPDLPROQVIRFKVACOOSRYIMISKDQITAKVYVLOAIRFAPVAT-PDQ	649
Db	605	---ISQYGVPRSCPEHVLAKTIRKNOQTRLYPKVKETSAQVAVO1ALQJENMFADESP	661
Qy	650	YSLCEVSTPEGVYKORRELPDQLSKADLRDILQSGRYLYKKNMETTLLCSDEDAQELLRES	709
Db	662	WSLCECJTIDGVYKQRLRPPQMEUAEIRLAINSRYYLTKNNSRSPPLVDLAELEKKEA	721
Qy	710	QISLLQSLTVEVATQLSMRNFELPRILEPTEYIDOLFUKRSTSCANKKREEVYINQET	769
Db	722	QJOLLSLNAOVYAAQJLQLOFSVFSALIEPTEFDNLFKDSYGSFKLEEEQJLNFEMW	781
Qy	770	VWASEILRETQOLRKMILIHFKIKIALHRECKNFEMSAIISGTLNLAIVARLRTTWEKL	829
Db	782	VWATEICTERHYQKRAKLTKIKFIVARCYGDLRNFENSMFAIMSGLDKPAVRRLSHSMERY	841
Qy	830	PNKTEKLEFQDLODLEPDSRMWAKYRNVLNSQNLQPIITLPEVIRKDDLTLFHEGSDVD	889
Db	842	SSKATIRMLDETHQLVDSRMSKSYRQHLVEAOEPPVYIYIKKDLFPAHDGATYSE	901
Qy	890	GLVNFELKRLIAKIEIRHYGVMA5YNDPALMFTRRKKRSLGSLSGOSTNATYLDVAQT	949
Db	902	KLINEKRLIAKIRGVMKLSSAPYEI5MAE-----RSGGVYMDALLHNSFENSNV	955
Qy	950	GGHKRRVRRSSFNAKLTYEDQAMARKVYOXL5NLELMEDEESLOTL5LOCEPATNPLPK	1009
Db	956	ATMRKMGSGKONOPRRKVVYEQALMVRKVS5YLEGLHVVYNDMELDMSYDIEPQVOTAKR	1015
Qy	1010	---NPEQKKRPVKS5PYAPRAGSOQKQNL5LPQOQPPRAHKI---NOGLOVPAYSILXP	1064
Db	1016	GANS5STANIRVR5P--TP5LSL5Q5G5ADQ5SR-----HRLLFNGTG---SIS5AGG	1065
Qy	1065	RKKYVVKDLPPEQIN5POALKLITL5E5BSLE5RHKHQAEDIT5NNA5SOL5SP5PP5SP55	1124
Db	1066	GSK-----F5G5E5FOAVOKML5LYQ5NK-----VK5AP5QIOT5P5ST5SAR55	1106
Qy	1125	-----PRKGYTLAP5-----GTVDFN--SD5G5H5ET5SR55ST5VNS5F5D5VY	1165
Db	1107	LQRMNPRVYTGROAT55AG5APQVLNEET5YTVTY5Q5D5GNRQ5RG-----SEGR5DNT5P	1161
Qy	1166	S---LHDER---QRH5V5T5ENTL5GMGRERERTM1LEPQOY5L5G5APM5B5EG5LYAT5	1219
Db	1162	5TFEYTL5G5LTV5PPO5L5SVY1P7PHPHG-----5PT5SPRC5R5P55SC55CF5TIA	1214
Qy	1220	VY55P5TEEL5Q5DQ5DRASL5DAD5G5RG5WT5C5SG5SHN1Q1Q-HQ5RW5ETLPE---G	1275
Db	1215	5TAA-----5T5MAA5S-----AV5N5PQ5HQ5TV5GHV5IGH5PM5YV5G	1255
Qy	1276	HTH5EY5G5DP5AGLMA55SHDQ1M5F5D5H5TKYKROQ5E5S	1316
Db	1256	SATL5ENHV5P5G5L5PK5--R5PTL1P5G5HT5SS5R5M5Q5T5IEA	1294

RESULT 2  
S42368  
quantine nucleotide releasing factor homolog - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*  
C:Date: 07-Oct-1994 #sequence\_revision 10-Nov-1995 #text\_change 16-Jul-1999  
C:Accession: S42368  
R:Smith, A.  
submitted to the EMBL Data Library, March 1994  
A:Reference number: S42368  
A:Accession: S42368  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1213 <SMI>  
A:Cross-References: EMBL:230423, NTD:q458479, PID:q458480  
C:Genetics:  
C:Introns: 196/3; 238/1; 263/3; 312/3; 444/3; 519/3; 550/2; 669/2; 742/1; 776/3; 1072/2  
C:Superfamily: CAM receptor protein cyclic nucleotide-binding domain homology <CDC25  
E:574-688/Domain: CAM receptor protein cyclic nucleotide-binding domain homology <CA  
E:987-1206/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match	7.08; Score 540; DB 2; Length 1213;
-------------	-------------------------------------

```

Best Local Similarity 22.4%; Pseud. NO. 1.4e-15;
Matches 231; Conservative 153; Mismatches 337; Indels 308; Gaps 35;

```

7 PANHGVMGOOEKHSIPADFTKLH-----LTDSLHPQVTHVSSSHS 46

388 PMNRYVYNOKNEE--RNETFEHNPTGTEKQIRDSGGTI-HRKMLTDN-HOYIBDTTEHT 444

```
A7 -----CCCCCCTTCCATTCESBACNDMDCSCIBETAVDSEDD--DNF Q1
```

[illegible]

1. The first step is to identify the problem or question that needs to be answered. This involves understanding the context and the specific requirements of the task.

27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100																										
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

505 NSYLRWVQDFRSRKNKVAAPSIEEVSKSITLLSSVAPEILFLMIVSNFGRFENRFEELVIE 504

129 FMHQLPFAFANMTMSVRRELCAVMFAVERAGTILVNDGEELDSWSVLNGSVEVYIPDG 188

565 ELTIFIKALSHLSTMKRQLSNFVKVEQYVHAGSVFRQGEIGVYWIYVLKGAVEVNV-NG 623

189 KAEILCM--GNSFGVSPMTDKKEYMKGVMTKVDDCQFVCIAQQDYCRILNQVEKNMKV 245

624 K-IVCLREGDDEFGKLLALVNDLPRAATIVTYEDDSMFLVVDKHHFNQILHQVEANTVRL 681

246 EEEGEIWMKEHRELDRTGTRK-----CHIVIKGTSERLTMLHVE-----EHS 288

682 K D Y G E D V L V L E K V D I P R G A L E N S N S C N F N C G Y S V M A G K A E I L E Y V L E T R I D A L G D D I S 741

289 VDDPTFIEDFLTYRTFLSSPMVEGKLLWFNDP--SLRDKVTRVLLWNHNFNDFEG 346

```

742 ELD-VFVEDEFLTHDAFMPDNTVCNFKSYFRTPTRYATRDSITD----- 785

```

347 DPAMTRLEEEENNLEREKMGHLRLNIACA--AKAKRRLMTLTKPSREAPLPFILGG 404

```

: | : | | : : | : :
-----SCTEEVRCJKRBY-----YOFVYYWC 805
786 -----

```

АВСТРАЛИЙСКО-КАНАДСКИЙ КОМПЛЕКСНЫЙ ПРОЕКТ

	:			:	617
	:			:	617

[illegible]

405 INLEVFNEELLIRKSEENKNGAPH--LPAIGDINAKSRISIFDRAVDVEQVIOLENNVNNNS 222

818 INSEVELFCHVIBDKRRLGGMEDLIRIGSLRST----- 832

523 KANTVGRNKLKILDKTRISILPQKPYNDIGIGSQDDSI VGLRQTKHIPTALPVSGTL 582

Db 853 -----RENMQLVL-----ARH-PAIVLDCGVL 873

583 SSSNPDLQSHHRLD FSATPDL PDQLR - VFKADQSR YIMISKDTAKEVIAQIRE 640

874 SAHTP-----CPVLPSDVCNQI IYLADTTCFVLP I RVDKTAEEICELSRRR 919

641 FAVTATPDQYSLCEVSVTPEGVI--KQRRLPDQLSKLADRIQLSGRYYLKNMETETLTC 697

```
920 MSFSAEP--LNLYEYKSNGEKLTFSPNDRAIPTVLS-----LNSKLYVVNRREIPLLV 970
```



```
QY 812 SGLNAPVARTTWEKLPNKYEKLFQDLQDLFDPSPRMARYRVNLQNLQPPILPFP 871
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 925 AAMGTAAYRRLWTFQGLPERYRKFLBECRELSDD--HLKTYQERLRISIN--PCVPFFG 980
QY 872 VIKKDLFTLHGNDKSVGC--LVNEFKRLMAKETRNHGRNASVMDALMFRIRKKWR 929
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 961 RYLNLHLLEGNDPLDLANTLNLINSKRKVAETII----- 1015
QY 930 SLGSLSQGSTNATVLDVQOTGCHKKRVRRSSFNAKLKLE-----DAQMARVQY 980
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1016 -GGLQYQONQPYCLNEST-----IROFEDLDFPGLNGLSDKQMS---DY 1055
QY 981 LSNLELENDSESLQTLQCEPATNTLPKNGDK-----KPVKSETSPVAPRAGS 1030
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1056 LYNESLRLEIPGCKTVPKFPKRWPHIPLKSPGIKPRRONQTNSSSKLSNSTSVAAAAA 1115
QY 1031 QOKKOSLPQPOQPPPAKINQGLQVPAVSLYPSRKVKPVVDLPFGINSQALKKLISL 1090
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1116 SSTATSI-----ATAPSLHASSINDAP----- 1139
QY 1091 SEEGSLERHKKQABDTISMASSQLSPPSPSPKRGYTLAP-----SGTVDN 1139
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1140 -----TAAANAGSGTLAGQSPQHNPFAVPAVILPERNTSSMSGTPQH 1186
QY 1140 F-SDSGHSEIS-----SRSSIVSNSPDSV-----PVSLHDER 1171
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1187 TRTOONGEVSVPAHPHLPKKGAHVMANNNSTLASASMDVFSAPALPEHLPPQSLPDSN 1246
QY 1172 RQRHSVSVETNLGCMERRTMIEPDQYSLGSAVPMSEGRGLYAFATVISPSPELSQ 1231
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1247 PFASTDTEAPRPL-----PLVYSP--RHETGNRSPF--HGRONSPHSTASTVTLT- 1295
QY 1232 DQGRASLDAADSGRGSKTSCSSGSHDNIOQTIOHRSWETLPFG-----HTHFYDSD 1284
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1296 -----CMSTSGGEFEFCAGGFYFN--SAHQGQPGAVPISPHVNVPMATMEXRAV 1342
QY 1285 PAGL-----WASSSHMDQMF-----DHST--KYNRONOS 1313
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1343 PPPLPPRRKETESCADMAQRCQAPDAPLPRLPDGELSPPIPLRLHSHGISYLRQSHG 1402
QY 1314 RESLEQAQSRASWASSTGYWGEDS--EGDTGTIKRGGKDVSIKFAESS-----S 1360
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1403 K-----SKEFGNSSLPLPNTSSIMIRNSAIEKAAATISQPNQAAAPIS 1448
QY 1361 LFTSTTEET---KVPMPAHIAVASTTYGLARKEGRYREPPRP-----PGYIGIP 1410
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1449 TTLVTVAQAVATDEPLPLPISPAASSSTTTS-----PLTPAMSPMSPNIPSHP 1496
QY 1411 ITDPEGHSPARKPPDYNNVALQSRMVARSSTAGPSSVQQRPHGHTSSRPVAKPQMK 1470
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1497 VESTSSSYAH-----QLRRKQOQOQOOTHPAIYSOHQHHTHLP--HHPHQH 1542
QY 1471 PNESDPRLAP 1480
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1543 SNPTQSRSP 1552
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
PC1114
SKDC23 protein - yeast (Saccharomyces kluyveri) (fragment)
C:Species: Saccharomyces kluyveri
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 21-Jul-2000
C:Accession: PC1114
R:Prigozy, T.; Gonzales, E.; Broek, D.
Gene 117, 67-72, 1992
A:title: Identification and analysis of a DNA fragment from Saccharomyces kluyveri that
A:Reference number: PC1114; MUID:92354938; PMID:1644315
A:Accession: PC1114
A:Molecule type: DNA
A:Residues: 1-1095 <PRT>
A:Cross-references: GB:M82964; NID:g171186; PIDN:AAA34479.1; PID:g171187
C:Genetics:
A:Gene: SKDC23
```

```
C:Superfamily: CDC25-type guanine nucleotide exchange activator homology
C:Keywords: Transmembrane protein
F:808-1049/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match          4.2%; Score 325.5; DB 2; Length 1095;
Best Local Similarity 21.6%; Pred. No. 1.1e-08;
Matches 158; Conservative 110; Mismatches 240; Indels 225; Gaps 27;

QY 434 LEVNGQNFENIQ-LSKMELLRNTHLSTIVKTNL--FYVKELLRLSEKRG-----A 485
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 380 LELSDATYKALNTYVDMEIEN---LDLTFVNLNMSHDEKLDRESELKRHAMTSLA 436
QY 486 PHLPIKIDIKKASRYSIPDLAV---DY-----EQTGLEKYNKSKAN 525
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 437 STLMEFEFEINQA-----LHDIAIRQIMQVQGLTLDDPEVFCATIDESVCYDRDELSKIQ 492
QY 526 TVGGRNKLKILDKTRISILPQK--PYNDIGIGSQDSDSYGLRQTHIPTALPVSGTSS 584
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 493 FL---KQEFABELRNLLISQDVEYNDLAFDA--DQLVREAFARF---SDIGNTMQ 542
QY 585 SNPDLLQSHRIIDFSATPDLPPQVLRVFKADQSS----- 619
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 543 NVEQLIERENILNYPARKMKSDITLALMKGEQKFEDEDFDMASSAGENENLDFGDAQ 602
QY 620 -----RYMI-----SKDTAKEVYIQAIR 640
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 603 NKSPSRDIPWYLDESEHYSILYDNKNIGKTEKALLEHLTHSQSLDPSNLAMLLTFS 662
QY 641 FAYT-----ATPPOYSICEVSVPEGY-----IKORLPQQL----- 672
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 663 IFTTGEFLQALVERYNL---YPEGVSYEEYNIWEEKQKPKIVVIMKTLFQSYWT 718
QY 673 -----SKLADRIQLS-----GRYVILNNMET---FTLCSDEDAQEL 705
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 719 PSYEPGLDYMIGFAQLAKQKISGADVLLSAIKGRLSKMKNLKNVPESINFSDDGSS 778
QY 706 -----LRESQISLQSLSTEVAQTQSMRNELEFRNIETPEYID 743
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 779 TTVQSSRSSVASPVGSSSTTFGRMRKLKLDISLDYAKQLTIKHSFLFYKISPECLD 838
QY 744 DLF--KLRSKTSANLKRPEEYVINOETFPVASTILETQOLKRMKLIKFKIALHCRC 801
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 839 RTWGNKYCNMGSKNTTEFISNSHLTNVSEFMIVKQTDIKRIQILOPFIWAAHCHEL 898
QY 802 KNFSMFAIISGLNLAPVALRTTWKELPKYKELFQDLODLFDPSPRMARYRVNLQNSN 861
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 899 NNSSLIATISALYSPYIRLKTMAVAPEYKLLLELNTLMDSKAKNIRYRLQSLKSG 958
QY 862 LQPIITLPPVYIKKDLFTLHGNDG---KYDGLVNEK---LMLAKETRHVGR----- 909
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 959 -DFPCVPFFGVYLDLTLFTANGNPDFLHNRNTVLVNGKRVRLLEILKEISVYQSRHYKL 1017
QY 910 -----MASVMDPALMFRTRKKKWRSLGSGSTNATVLDVAQVGHGKK 954
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1018 RYEDIOAFIESSLENLPSTIEKOYASLRNEPREVSTG---INSTVNVYNTKNGPENR 1074
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 955 RV-RSSSFLNAKK 966
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1075 NTGKRLKFGKAKK 1087
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
S64758
SCD25 protein (version 2) - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein L1309; protein YL016w
C:Species: Saccharomyces cerevisiae
C>Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 21-Jul-2000
C:Accession: S64758; S64764; S69390; S70559
R:Miosga, T.; Zimmermann, F.K.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64743
A:Accession: S64758
A:Molecule type: DNA
```

A:Residues: 1-1048 <MIO>  
A:Cross-references: EMBL:273121; NID:g1360186; PID:e245452; PID:g1360187; MIPS:YLL016w  
A:Experimental source: strain S288C  
R:Goifeau, A.; Purnelle, B.  
Submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64761  
A:Accession: S64764  
A:Molecule type: DNA  
A:Residues: 1-1048 <GOF>  
A:Cross-references: EMBL:273121; NID:g1360186; PID:e245452; PID:g1360187; MIPS:YLL016w  
A:Experimental source: strain S288C  
R:Purnelle, B.; Goifeau, A.  
Submitted to the EMBL Data Library, April 1996  
A:Description: The sequence of 32 kb on the left arm of yeast chromosome XII reveals 14 mly and a new ABC transporter homologous to the human multidrug resistance protein.  
A:Reference number: S69380  
A:Accession: S69390  
A:Molecule type: DNA  
A:Residues: 1-1048 <PUR>  
A:Cross-references: EMBL:X97560; NID:g1297003; PID:e238680; PID:g1297014  
R:Mioza, T.; Zimmermann, F.K.  
Yeast 12, 693-708, 1996  
A:Title: Sequence analysis of the CEN12 region of *Saccharomyces cerevisiae* on a 43.7 kb e conductance regulator protein CTR.  
A:Reference number: S70557; MUID:96405918; PMID:8610043  
A:Accession: S70559  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1048 <MIM>  
A:Cross-references: EMBL:X91488; NID:g1495203; PID:CAA62775.1; PID:g1495207  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1995  
C:Genetics:  
A:Gene: SGD:SDC25; SDC25  
A:Cross-references: SGD:S0003939; MIPS:YLL016w  
A:Map position: 12L  
C:Superfamily: CDC25-type guanine nucleotide exchange activator homology  
F:744-995/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>  
  
Query Match 4.28; Score 321.5; DB 2; Length 1048;  
Best Local Similarity 19.98; Pred. No. 1.7e-08;  
Matches 226; Conservative 166; Mismatches 402; Indels 339; Gaps 45;  
  
QY 13 MGOEKHSLEADFTKLHLTDS--LHPQVTVSSSHSGCSTSDSGSSLSLDIYQATESEA 70  
DB 175 LNHQMSMDLFRQMKAVAGSAGIYAENIDELPASKQSTSCSSSETHSPAPQRRRGRT 234  
QY 71 GMDLGLPETAVDSEDDDEEDIERASDPLMSRDIVRDLCKDPIDRTDDIEQLLEFM 130  
DB 235 IFSNVSG-----SSDEPTIMSKRKRPYPLNEETL--SLVRARKKQLDGLKQIMKSA 285  
QY 131 HOLPAPFNMVSVARELCAMVAVRVR--AGTYLNDGEELDSKSYTLN-----G 179  
DB 286 NE--YLSNANFSKMLNFEMNFETVEVSGTPIIDILLENLDITFLNRELGDENRVF 342  
QY 180 SVEVTPDGKAEIL-----CMGNSFGVSPMDKEVKKGVWRKTVDDCOFACINQO- 229  
DB 343 DEVAIDDEDEEFLKHSLSISLISDYFMN-----KQYHADV-----VKFLIVAOHL 391  
QY 230 -----DYCRILNQVEKNMQKVEEGEIIYVWKEHRELDRTGTRKHGHI 272  
DB 392 TLDEPVPFSQNDLPRTGYEPMKPSNLINDMKDK----- 427  
QY 273 KGTSEKRLTMHLVEHSAVDPTFIEDPLLTTRTFLSPMEVKGKILKEMFNDSLRDKATRV 332  
DB 428 KNGSQNDIDEEDEYEPDP-----DSLILFNLINQDSDFND--LKRFFN-----LAHV 474  
QY 333 VLLVNHFPNDFEGDPMATRFLEEFENN--LEREKMGHILRLINIAACAARRLMTLTK 390  
DB 475 FKSKCDYFD-----VLKALIEFVQMLILER-----NLMTYA-----ARMKNITTE 517  
QY 391 PSREAPLPFLLLGSEKGFIFVDSVSGSKATEAGLRGQDILVNGONFENITOLSKAM 450  
DB 518 -----LLLRG-EEGYG-----SYDGGEMAEKS----- 538

QY 451 EILRNNTSLITVKNLFEVFKELLTRLSEKRNQAPHLPKIGDIKKARSYIPDLADVE 510  
DB 539 -----DTNA-VYADSDTKNDENRDSQVKLP-----RT----- 565  
QY 511 QVIGLEKVRNKRKANTV--GGRNKLKLLDKTRISILPQKPYNDIGIOSODDSVGLRQT 569  
DB 566 -----LQREYSELIMGWNNRIRKGSKHALLSYLDNDNKKDLFF-----DITFLITF 612  
QY 570 KHIPALP-VSGTLSSNP-----LLOSHRILDFSAT----- 602  
DB 613 RSIFTTTEFLSYLISQYNLDPPEDLCEEYNEWTKLIPVKCQVWEIMTFEQQYWEFG 672  
QY 603 PDDPDQVRFKADQOSRIYIMISKDTAKERVYQAIREFAVTAPPDQYSLCEVSPPEGV 662  
DB 673 YDEPD--LATLNDYFQVAKIKENTGVELLEVNOKF-----KHGN 713  
QY 663 IKORRLPDQLSKLADRIQLSGRYLKNMME--ETLCSDEDAQELRESOISLPLSTVE 720  
DB 714 IQENTAP-----MTLDQIQIDHISGLTSTTE-SILAVDPVL 751  
QY 721 VAFQLSRNELEFRNIEPTFYIDLEK--LRSKTSCANLKRFEVINQETFWVASELR 777  
DB 752 FAFQTLTEHEIYCEITIFPOLQIKWKNKYTKSYGASPLGMEFISFANKLITNFTSYSVK 811  
QY 778 ETNOLKRMKTIIRKIKALICRECKNNSMFAISGYNLAPVAVALTWTEKLPKYEKL 837  
DB 812 EADSKRAKLSHFIFAEYCRKNFSSMTAIIISALYSPIYLETWAVIYQTDLL 871  
QY 838 QDLQDLDPSPRNMAKYRNVLNNSONLQPIIPLPFVIKDLTFLIEGNSKVDGLVNEKL 897  
DB 872 QSLNKLMDPKPKNFINTYNELKSLH-SAPCVPFVEYVLSDLFTFDSGNP---DIVYLEHGL 927  
QY 898 RMIAKETIRHVGMAVSNMDDPALMFRTRKKWRSLGSLSGSTNATVLDVAQTGCHKRRVR 957  
DB 928 KGVHDEKKYIN-----FNKRSR-----LVDIQ----- 950  
QY 958 RSSSLNKKKILYEDQAQMARKKQYLSN--LELEMBESIQTSLSCEPRTNLPKPGCKRP 1016  
DB 951 -ETIYFKTHYDPTKDRTYECSISLENIPIHEIKOYQSLIIEP-----KP 996  
QY 1017 VKSETPVAPRAGSOAKOSLPQPOOPPPAHKINGQIQAQVAVSLYPSRRKVP 1069  
DB 997 RKR-----VVRNSNSNNKSQEKSRDDQTDGKTSIKKD-REPFRQLHKTKKKAP 1044  
  
RESULT 7  
S1417  
SCD25 protein (version 1) - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: protein L1509; protein YLL016w  
C:Species: *Saccharomyces cerevisiae*  
C:Date: 02-Jun-1994 #sequence\_revision 02-Jun-1994 #text\_change 06-Feb-1998  
C:Accession: S14177; S12942; PS0040  
R:Damak, F.; Boy-Marcotte, E.; le-Roscouet, D.; Guilbaud, R.; Jacquet, M.  
Submitted to the EMBL Data Library, August 1989  
A:Reference number: S14177  
A:Accession: S14177  
A:Molecule type: DNA  
A:Residues: 1-1250 <DAMI>  
A:Cross-references: EMBL:M26647  
R:Damak, F.; Boy-Marcotte, E.; le-Roscouet, D.; Guilbaud, R.; Jacquet, M.  
Mol. Cell. Biol. 11, 202-212, 1991  
A:Title: SDC25, a CDC25-like gene, which contains a RAS-activating domain and is a di  
A:Reference number: S12942; MUID:91094833; PMID:1986220  
A:Accession: S12942  
A:Molecule type: DNA  
A:Residues: 1-737, 'A', 738-970, 'I', 972-1250 <DAMI>  
A:Cross-references: EMBL:M26647  
A:Note: the authors translated the codon GAG for residue 538 as Asp, GTC for residue  
R:Boy-Marcotte, E.; Damak, F.; Camonis, J.; Garreau, H.; Jacquet, M.  
Gene 77, 21-30, 1989  
A:Title: The C-terminal part of a gene partially homologous to CDC25 gene suppresses  
A:Reference number: PS0041; MUID:89306677; PMID:2545538

A:Accession: E50404  
A:Molecule type: DNA  
A:Residues: 668-679, 'NPVMTIMC', 689, 'N', 691-1250 <BOY>  
A:Note: the authors translated the codon GCA for residue 747 as Asp and GAT for residue 748  
C:Genetics:  
A:Gene: SGD:SCD25: SPC25  
A:Cross-references: SGD:S0003939; MIPS:YLL016w  
A:Map position: 12L  
C:Superfamily: CDC25-type guanine nucleotide exchange activator homology  
F:946-1197/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 4.0%; Score 310.5; DB 2; Length 1250;  
Best local similarity 26.1%; Pred. No. 8.3e-08;  
Matches 105; Conservative 65; Mismatches 159; Indels 73; Gaps 12;

QY 680 QLSGRYYLKNMNET-----ETLCSDEDAQELLRESQISLLOLSTVEVATPQLSMRNF 731  
D 680 EVNQKFKNGNIOEATAPKKTLDQQLCDHYSGLTSTTE-SILAVDPVLFATQLTILEHE 964  
QY 732 LFRNIEPTVEYIDDLFK--LRSKTSKANLKRFEVINDETFWVASEILRETQOLRMKII 788  
D 732 IYCEITTFDCLQIKMKNYKTSYGASPLNEFISFANKLTNFISSVYKLEADKSKRAKLL 1024  
QY 789 KHEFKIALHCECKNENMFALISGLNAPYARLRTTEKLPNKYTEKLPQDLQDLDPSS 848  
D 1025 SHEFTIAYCYCKNENFSSMTDIIISALYSSPIYRLEKTMQAVIPQTRDLOSLINKLMDPK 1084  
QY 849 NMAKYRNVNLSNQLPPIPLFPIYIKKDLTFLEHGNSKVDGLNFEKRLRIAKEIRHVG 908  
D 1085 NFIVNRMLKSLH-SAPCVPFEGVYLSDLTFTDSGNP---DYLVDEHGLKGVHEKKIYN 1140  
QY 909 RMAVNMDDPALMEFRTRKKKKRSLGSLSGSTNATVLDVAQTGCHKRVRSSFLNARKLY 968  
D 1141 -----FKRRSR-----LVLDLQ-----EIIYFKKTH 1161  
QY 969 EDAMARKVQYLSN-LELEMEDESLOTLSQCEPATNTLPKNPGDKPKVSESPVAPR 1027  
D 1162 YDFKDKRTVIECISNLENTPIHEKQYQSLITRP-----KPRKK-----VVPN 1205  
QY 1028 AGSOQKAQSLPQPPQPPRAHKINQGLQVPAVSLYPSRRKVP 1069  
D 1206 SNSNKSQEKSRDDQDTDEGKTSTKKD-RFRPKQJHKTKKKAP 1246

RESULT 8  
S28407  
guanine nucleotide-exchange activator CDC25 homolog - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 17-Apr-1993 #sequence.revision 17-Apr-1993 #text.change 05-Nov-1999  
C:Accession: S28407; S22693; B46199; S20730  
R:Gen, H.; Papageorge, A.G.; Zippel, R.; Lowy, D.R.; Zhang, K.  
EMBO J. 11, 4007-4015, 1992  
A:title: Isolation of multiple mouse cDNAs with coding homology to Saccharomyces cerevisiae  
A:Reference number: S28407; MUID:93010996; PMID:1396590  
A:Accession: S28407  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-1260 <CEN>  
R:Marlegani, E.; Vannoni, M.; Zippel, R.; Cocciotti, P.; Brambilla, R.; Ferrari, C.; Stura  
EMBO J. 11, 2151-2157, 1992  
A:title: Cloning by functional complementation of a mouse cDNA encoding a homologue of Cdc25  
A:Reference number: S22693; MUID:92289680; PMID:1376246  
A:Accession: S22693  
A:Molecule type: mRNA  
A:Residues: 789-1260 <NAR>  
A:Cross-references: EMBL:X59868; NID:950357; PIDN:CAA42525.1; PID:950358  
R:Wiel, W.; Mosteller, R.D.; Sanyal, P.; Gonzales, E.; McIntney, D.; Dasgupta, C.; Li, P.  
Proc. Natl. Acad. Sci. U.S.A. 89, 7100-7104, 1992  
A:title: Identification of a mammalian gene structurally and functionally related to the yeast Cdc25  
A:Accession: B46199; MUID:92357779; PMID:1379731  
A:Status: preliminary  
A:Molecule type: nucleic acid

A:Residues: 1029-1030, 'D', 1032-1224 <WEI>  
A:Experimental source: fetus  
A:Note: sequence extracted from NCBI backbone (NCBIN:111101, NCBI:111102)  
C:Superfamily: CDC25-type guanine nucleotide exchange activator homology; CDC24 homol  
F:242-428/Domain: CDC24 homology <CD24>  
F:1021-1257/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 4.0%; Score 309; DB 2; Length 1260;  
Best local similarity 19.8%; Pred. No. 1e-07;  
Matches 219; Conservative 157; Mismatches 364; Indels 364; Gaps 44;

QY 17 EKHSLPADFTKL-HLTDSLHPQVTHVSSN-----SGCSITSDSGSS----- 57  
D 407 ERNSLDYAKSKLEELSRIMHDEVESETENIRKNLAIERMTGCGTLLDTSGTFVROGSLM 466  
QY 58 -SLSIDYQATEESAGMDLSGLPETVAVSDEDDDEEDIERASDPLMSRDIV----- 107  
D 467 QMSLSEKSKSRGLGSI-----STKKEGROCF-LPSKHLIITRSGSG 510  
QY 108 -----RDCLEKDPIDRTD-----DIEQL-----LEFNHOLAFANMTMSV 143  
D 511 KLHETKGVISLIDCTLLDEPENLDDEAKGAPETHELEKIGVEPKDSLPTVYIVAST 570  
QY 144 RRELCAVM--VFAVVE--RAGTIVLNDGEELDSWVILNCSVEVTPD--GRAELICMG 196  
D 571 ROEKAAMTSDIICQVDNIRCGMLMNAFEE-----NSKVTVPQMTKSDASLYCDD 620  
QY 197 NSFEVSTPMDEKRYKGVYKRTKVDCCQVCIAODYCIILNOVEKNMKQVEEGEIVYVKE 256  
D 621 VDIREFSKTNMS-----CKVL-QIR----- 638  
QY 257 HRELDRTGRKGIIVIKGTSERLTMHLEHESVVDPFIDFLITLTPFLSPMEYCKL 316  
D 639 -----YASVERLLERLL-----DLRFLSIDFLMTFLHSTRVF--TNAMVYLDKL 660  
QY 317 LEWNPDSLQDKYRVVLLVNNH-----FNDEGPDAWTRFLEEDENNLERKMGHLR 371  
D 681 INIYRKPMSAIPARSLLELFSSSHNAKLYLGDAKPSRASR--KSSPP----- 728  
QY 372 LNTACAKAKRRLMTLTTPSRREAPLPFIILGSEKGFIFVDSVDSGKATAGLRGD 431  
D 729 -LAIGTSSPSRRKLSLNIP-----ITGKALFLA----- 758  
QY 432 QILEVNGQNFENIOLSKAMEILRNTHLSI-----TVKNTLFFVEKELLRLSEKRNQA 485  
D 759 -SLGSSDSYANI-----HSPISPGKTTLDGKLCMASSLPKPTPEELDVYA 804  
QY 486 PHLPKIDIKKASRYSLPDLAVDEOYIGLEKYNKSKANT---VGRNKLKLIIDKTR 541  
D 805 TIPEKPELSSAKRHSDDVLKEESE-----DDONHSDENITEVSPKSPPTPKSFLNRT- 858  
QY 542 ISILPQKPYNDIGIGQSDSIYGLQTKHIIPALVYSGTLSSNPDLDQSHRILDFSA 601  
D 859 IIEPFPFNNGILMTTCRDLVNNSTLSATSAFAPATAGAGESNKEVFRR-NSLN 917  
QY 602 TPDLDPQVLEVRKADQS--RYTIMISKDTAKREVIVQAIREFAVTAPPOYSLCEVSATP 659  
D 918 TG-----FSSDQRNIDKEVIRAAATNR--VLNLRHWTKRTQD-----PDT 958  
QY 660 EGYIKORRL-----PDQSLKADRIQLSGRYLYKNNMETELTCSDEDAQELLRESQ 710  
D 959 DDLTKRVICFELEVNHDPDLTQ-----ERKAANIIRTLLEETEQHSMLEEVY 1010  
QY 711 ISLQLOST-----VEVATOLSMRNFELRNI-----EPREYIDDLF 746  
D 1011 LMEGYKTFEENHPRLDIAEQLTLLDHLVFSIPYEFPFGQGMKAKEYERTPT- 1066  
QY 747 KLRKTSKANLKRFEVINDETFWVASEILRETQOLRMKIIKHFKIALHCECKNENS 806  
D 1067 ---MKTT---KHFNVSN---FISSEIIRNEDISARSALEKWAIVADICRCHLNHYA 1115  
QY 807 MFAIISGLNAPYARLRTTEKLPNKYTEKLPQDLQDLDPSSRMKAYRVNLS-ONTOPP 865

Db 1116 VLEITSSIRNSAIFRLKKTWLVKSQTKSLDLKLOKLVSSD---GFKULRESLRNCDDP 1172

QY 866 IIPLEFVIKDLTFLEGNDSKV-DGLVNFELKRMIAKEIRHV--RMASVNMDDPALMR 922

Db 1173 CVPYLGMYLTLVLEIEGFPNTEDGLVFMKRMISHIIRERQOQTYTIDP----- 1227

QY 923 TRKKWRSLGSLSGSTNATVLDVAQTGGHKRRVRRSSFLNAKKLYEDAQMAKKYKQYLS 982

Db 1228 -----QPKVIQYLL 1236

QY 983 NLELMDSESLDTLSLOCEPATNT 1006

Db 1237 DESFMLEDSLEYESSLLIEPKLPT 1260

## RESULT 9

Ras guanine nucleotide exchange factor son-of-sevenless (sos) 1 - mouse  
 S25716  
 A:Residues: 1-1336 <BOM>  
 A:Cross-References: EMBL:211574; NID:954134; PIDN:CAA77662.1; PID:954135  
 C:Species: Mus musculus (house mouse)  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 05-Nov-1999  
 C:Accession: S25716; S21391  
 R:Bowtell, D.; Fu, P.; Simon, M.; Senior, P.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 6511-6515, 1992  
 A:Title: Identification of murine homologues of the Drosophila son of sevenless gene: pc  
 A:Reference number: S25714; MUID:9235328; PMID:1631150  
 A:Accession: S25716  
 A:Molecule type: mRNA

Query Match 4.0%; Score 307.5; DB 2; Length 1336;  
 Best Local Similarity 19.6%; Pred. No. 1.3e-07;

Matches 224; Conservative 169; Mismatches 409; Indels 339; Gaps 47;

QY 236 NOYKMKQVEEGEL-VWVKEHRE--LDRTGRKGHIYIKGTSERLTVMHVEHGVDP 292

Db 257 NDVENFISIVDHELHLSVLLGHIEPTVEMTDEGSPHPLVSGCEDLA-----ELALFDP 311

QY 293 --TFIEDFLLT--YRTFLSSPMEVGRKLL-----EMFNDPSSLRDKVTRVLLVMNNHFPD 343

Db 312 YESYARDILRPGHGFHLSQLSKPGALYLSIGEGFKE-AVQYVLRLLAVNYCLHY 370

QY 344 FEGDPMATFLFEFENNLREKMGHL-RLN-----IACAARAKRLMLTLTKPSREA 395

Db 371 FE--LLKOLEKSEDOEDKECKQAITALLVQSGMERIKCSLSLAKRRL----- 417

QY 396 PLPFIILGSEKQFGLFVSDVSGSKATEGLRGDQILEVNGQNFENIOLSKAM-ELIR 454

Db 418 -----SESCRYTSOOM-----KGKOLAIKKMMETIQKIDGMEGKDIOGCCNEFTM 463

QY 455 NNTHLSTVTK--TNLFVKELLTRISEEKRNGAPHLR-----KIGDI 494

Db 464 EGIITFVGAKEHNRHIFLPGLM--ICCKSNHGQRLPGASSAEYRLKEKFPMKVQINDK 521

QY 495 KASAR--SIRPLAVDEQYI-----GLKVKKKSKANTVGGKNAKKLLDKTRISILQ 547

Db 522 DDTSEKNAFEIILKDGNSVIFSASAEKNNMMALISLQYRSTLERMD----- 572

QY 548 KPVNDIGIGOSODSIVGRQTKHIFTALPVSGTSSNPDDLQSHNRI--DESATPDL 606

Db 573 -----VTVLOEKEEOMRLPSAVYRFAERD--SEENILFEENVPKKG 614

QY 607 DOVLK--VFKADQOSRYIMISKDTAKEVYQALIREFAVTATPDQSYLCEVSVTPREGVI 663

Db 615 IPIKAGVTKLIERLTGMHYADP-----NFVTRFLTT--YRSCRPQELLSLI 662

QY 664 KQRLPDLQSLADRIQLSG-----RYLKNMNETFLGS-----D 699

Db 663 ERREIPEPEPTADRIAIENGDPQLSAELKRFREKEYIQPVOLRVLVNCRHWENHNFYDE 722

QY 700 EDAQELLRESQ----- 710

Db 723 RADDLQRRHEEFGTGRGAMKKWVESITKIIQKRIADNGCHNITQSSPTVEWHI 782

QY 711 -----ISLQSTVEAVATQLSMRNFELFRNIEPEYIDDLFKLRK--TSCANLKREE 761

Db 783 SRPGHIEFDLLTLPRIELARQLTLESIDLVRAGVSELVSGVWTKEDKEINSPLNKKM 842

QY 762 EVINQETFWVASIILETQOLKRMKIIKHEIKALACRECKNNSMFAIISGLNAPVAR 821

Db 843 RHTNTLTMFEKICIVETENLEERVAVVSRIELIQLVFOELNNNGVLEAVSANNSSPYR 902

QY 822 LRTTMEKLPNKYEKLFODLQDLFDP--SRNMAKYRWLVNSQNQLPIPLFPYIKKDLTF 879

Db 903 LDHTEFOIPSRQKKILEEAHELSEDIHKYTLAKRSI-----NPPCVPEFGIYLTNLIK 956

QY 880 LHEGNSKYDGLVNFELKRMIAKEIRHVGRMASVNMDDPALMFPTRRKKRSLGSLSGST 939

Db 957 TEEGNP-----EVLRRHGKEL-----INFSKRRRVAEITGEIQOYN 993

QY 940 NATVLDVAQTGGHKRRVRRSSFLNAKKLYEDA-----QAKRYKQYLSNLELMDSESLQ 994

QY 995 TSLJOCEPA-TNTLPKNPQDK--PVKSE-TSPVAPRAGSQKQASLPQPOQOPPPAAH 1049

Db 1034 -----EPRAHPKLPKRP--KTSYLPKSGVPRSPNREOTMH---PTPLOEP--- 1077

QY 1050 INQGLQVPNAVSLVSRKKVYPKVDLPFGINSQALKKIILSSEGLSERKKRQAEPTISN 1109

Db 1078 -----RKISYSRIIP-----ESTETESYASA 1096

QY 1110 ASS--QLSPPTSPQSSPRKGTTLAPSGTVDNFSDSGHSEISRSISYNSNDFDVPYS 1166

Db 1097 PNPSPRTPLTPPPASGTSMTDWCSPDS--DHSASPFRHSASVSISLSKGTDEVPVP 1153

QY 1167 LHDERQRHSVSIVETN---LGMGRMERRTMIEPDQYSLGYSVPMSEGRGLVYATAVIS 1223

Db 1154 PVPYPPRRPESAPAESPSKIMSKHLDSPATPRPPTSKAYSP---KISIDRTISID 1209

QY 1224 P 1224

Db 1210 P 1210

## RESULT 10

A37488

Ras guanine nucleotide exchange factor son-of-sevenless (sos) 1 - human

C:Species: Homo sapiens (man)

C:Date: 03-Mar-1994 #sequence\_revision 18-Nov-1994 #text\_change 18-Mar-1997

C:Accession: A37488

R:Charadri, P.; Camonis, J.H.; Gale, N.W.; van Aelst, L.; Schlessinger, J.; Wigler, M.

Science 260, 1338-1343, 1993

A:Title: Human Sos1: a guanine nucleotide exchange factor for Ras that binds to GRB2.

A:Reference number: A37488; MUID:93262494; PMID:8493579

A:Accession: A37488

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1333 <CH>

A:Note: sequence extracted from NCBI backbone (NCBIP:132148)

C:Superfamily: CDC25-type guanine nucleotide exchange activator homology; pleckstrin

F:442-544/Domain: pleckstrin repeat homology <PLK>

F:776-1019/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 3.9%; Score 303.5; DB 2; Length 1333;

Best Local Similarity 19.0%; Pred. No. 2.1e-07;

Matches 221; Conservative 175; Mismatches 397; Indels 371; Gaps 47;

QY 236 NOYKMKQVEEGEL-VWVKEHRE--LDRTGRKGHIYIKGTSERLTVMHVEHGVDP 286

Db 240 NDVENFISIVDHELHLSVLLGHIEPTVEMTDEGSPHPLVSGCEDLAELADPYESYA 299

Qy	287	HSVDPFIEDILLTYTFTLSSP-----MEVKKLLEFNPNDPSLRDQVTRVLYLWVNNH	340
Db	300	ROLLRPGHDFL-----SOLSKPALYLQSIGEGFKE-----AQYVLPRLILAPYHC	350
Qy	341	FNDEGDPAMTRLEEFENLREKMGHIL-RLN-----IACAARAKRLMTLKPS	392
Db	351	LHFE-----LLKQLEKEDEDECKLOKATITALLVQSGMEKICSKSLKRL-----	400
Qy	393	REALPFLILGSGSKGCIIVDSVDSGSKATEAGIKRGOLILEVNGQNPENIOLSKM-E	451
Db	401	-----SESACRYSOOM-----KGOLAKMMEL-OKNIDGEGKDIGCCNE	443
Qy	452	ILRNNTHLSTYVK--TWLFYKBEILTRLSEKNGAPHL-----KI	491
Db	444	FIMEGTILTRGAKHERHIFLDGIM-ICCKSNHGPRILGASNAEYRLKEKFEFMRKQYI	501
Qy	492	GDIKASKRY--SIDPLADVQYI-----GLEVNNKKSANTVGGKRNKLKTLDKTRSI	544
Db	502	NDKDDTMEYKAHEFILLKDNESVIFSAKSAEKNMMMAILSYRSLTERLIDVTM--	558
Qy	545	LPOKPYNDIGIGQSDDISIGLROTKHIALPVRGSLTSSNPDLLOSHNRILRDSAPD	604
Db	559	-----LOEKEEQRMLRPSAVYFAERD--SEBNIIIFEENMPK	595
Qy	605	LPDQVL--VEKADQOSRYIMTSKDTTAKEVYQAIREPVA7ATPDQYSLCEVSPTEG	661
Db	596	AGIPRLKGVYIKLIERLFTYHMYADP-----NFRFTFTY--YKSFCKPQELSL	643
Qy	662	VIKORRLPDQSLKADRILOSG-----RY-----	685
Db	644	IIEIRELPERPTEADRIALBNGDPLSALFKFKEFYIQPOLKVLNCRIMVWHNEYD	703
Qy	686	-----YLNKNNH-----TELCSDEDAOELLRS-----	709
Db	704	FERDAYVLLORHEEFTGVRGKAMKMWESTYIKLQKKKILAPRNGHNITFOSSPTEW	763
Qy	710	-----QSLLOSTVEVATOLSMRNEFLRNPTEPIDDLFLKLSK--TSCANLKR	759
Db	764	HISRGFIETFDLLTRPIELARLITLESDDLRAVQPSBLVGSVWTAKDKKITSNPILIK	823
Qy	760	FEEVINDQTFWVASEILRETNOILRMKILNHFIKIALHGRECKNFNSMFAITSGNLAPV	819
Db	824	MIRHTNTLTFEKCIVETENLEERVAVVSRIEILQVFOELNENGVLEVANSNSPV	883
Qy	820	ARLRTTKLRLKRYEKLFPDLODLFDP--SRNNAKTRANVLNSONLOPRLIPEFYIKDL	877
Db	884	YRLDTEFQILPSROKKLILEENHEISEBHYKYVLAKRSI-----NPPCVREFGYITNI	937
Qy	878	TELHGNSKSDVGLVNEFKLIMLAKELIRHGYMASVNDPMLMFTYRKKKWMSLSLSQG	937
Db	938	LKTEGMP-----EVLKRHKEL-----INFSRRKVALITELIOY	974
Qy	938	STNATVLDVAOTGGHKRRVRRSSFLNAKLLEYDA-----OMARKYKQYLSNLEMEDEBS	992
Db	975	QNOPCYLAVES-----DIKREFENLNMGNSMEKEFYDYLFNKSLIEEPNN	1020
Qy	993	IQTSLSGCEPRTNLLPKNPGDK--PYKSE-TSPVAPRAGSOQKQOSLPDQOQPPAH	1048
Db	1021	-----PKPLPRP--KKTSTYLLKSPGYRPSNPRGTGRH--FTLQOEP--	1066
Qy	1049	KINQGLQVPAVSLYPSRKRYVNDKLPFGINSQALUKILSLSEEGSLERHAKQOAEITS	1108
Db	1061	-----RKISYSRIP-----ESMETAS	1078
Qy	1109	NASS--QLSSPPISSPOSSPKKGTTLAPSGVDNFS--DSGHS-----	1146
Db	1079	AVNSRTRPLTPPPASGAS-----TTDVCYSVDSHSSPHSSNDTVFTQVTLR	1127
Qy	1147	---EISSRSSIYVNSSFFSVSLDHERQHRSHSVIETN--LGMGRERRTMILEPDQY	1200
Db	1128	HQPRASVYSSTSLTKGTDENVVPPVPRRRPRPSAPRESSPSKIMSKHLDSPALPRQP	1187
Qy	1201	SLGSTATPMSSEGRGLYATVATVISP	1224

[illegible]



Db 921 SLDSNENA-----VF-----LLMFSSMMULGELLISLLARFNIE----- 954  
QY 654 EYSVPEGY-----IKORLPDQSLK-----A 676  
Db 955 ----PPEGLSYEEYMLWMSKKRNPIRLRVINIMKLLLEKNWSMSYNEPVLRLMTFAHS 1010  
QY 677 DRIOLSGR-----LYLKNMETETLCSDDED-----AQELLRESOISLLOS 717  
Db 1011 DOVOYITSLGNLNVNLERLGERIRIVERDPVPIPNTPKPAPLTKGSSLLKPKPV--MDID 1068  
QY 718 TVEVATQISMRFNEFERNIEPTETIDLFKLS--KTSKANIKREBEVINOETFWVASEI 775  
Db 1069 YVELARQLTLREFKLYCKITFKACLAKVWGKSGLSSESIDSTQFIKASNOQLTFVGYMI 1128  
QY 776 LMETNOLKMKKIIKFIKIALHCRECKNNSMFATISGLNLAVALRTTWELPKYK 835  
Db 1129 LRRADPKRRVQIIRFYIYADCKROYNFSMATAISALYSPIHLLKTKTWEWMNADALS 1188  
QY 836 LFQDLODLDPSPRNMAKYNNVLNSQNLQPIIPLPFPVKKDLTFLEHGNDSKV---DGLV 892  
Db 1189 NLKNNKMLNNSRNREYEDVLKFTGSE--PCVPEFGVYISDLTFVYHGNPDLYNTRQY 1247  
QY 893 NFEKLMIAKEIRHNGR-----MASVNDPALM 920  
Db 1248 NFAKRAKTSSEIYSGIDRFKTTGYNFOEVEPEIQKFLDAMFEKCPPTIDEQYQISLNEP--- 1304  
QY 921 FTRKKKMSLSLSGSGTNAT 942  
Db 1305 -----REOPAGASNSNSTNAT 1321

## RESULT 12

JC7736  
C3G protein, long type - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 14-Dec-2001  
C:Accession: JC7736  
R.Zhai, B.; Huo, H.; Liao, K.  
Biochem. Biophys. Res. Commun. 286, 61-66, 2001  
A:Title: C3G, a guanine nucleotide exchange factor bound to adapter molecule c-Crk, has  
A:Reference number: JC7736; MUID:21378139; PMID:11485308  
A:Contents: 373-L1 adipocyte  
A:Accession: JC7736  
A:Molecule type: mRNA  
A:Residues: 1-1086 <ZHA>  
A:Cross-references: GB:AF348669  
A:Comment: This protein, a guanine nucleotide exchange factor, through the interaction  
ation.

Query Match 3.8%; Score 295; DB 2; Length 1086;  
Best Local Similarity 23.5%; Pred. No. 4,1e-07;  
Matches 150; Conservative 82; Mismatches 205; Indels 202; Gaps 26;  
QY 491 IDIDKASRYSPDLAVDEVOYIGL-EKYNKSKANTVGGRNKLKILDKTRISILPORK 549  
Db 535 VGD-----FSYVELAGDTEKPPLEPKKNKMLAYM-----QLLDYSPQPSMFTYQTP 583  
QY 550 YNDIGIGGSG-----DDSIVGLRQTKHT--PTALP----- 577  
Db 584 QSEHIVQOKNKMMLMEYVGFSEFCGSDSGSTOELAPRALPPKORLEPPSGKDGHPDPDV 643  
QY 578 --VSGTLSSSN-----PDLQ 591  
Db 644 SSSSGKDSNENGERSPKSLDGLESAQSEBEVDLSLDHNEIMARTLTLOEGDGDGVNRG 703  
QY 592 SHHRIIDFSAT-PDLPDQVL-----RVFKADQ-----QSRVIMIS--KDTTAKEV 633  
Db 704 GSGDILLVHATEDRKDLVLYCEAFITTYRTFISPELIIKKIQRYREKSPFADTKKRV 763  
QY 634 VIOAIEPAVTATPDQYSLCEVSVTPPEGVIKORRLPDQSLKADRIQLSGRIYLKNNMET 693  
Db 764 SKNTF--FVLVRVDE--LCLVELTEE-ILK-----LIMEIVRLVCSGSELSTARVLR- 811

QY 694 ETLCSDEDAOELLRESQI-----SLQSLTVEVATQISMRFNEPT 739  
Db 812 KNILDKVDQKLLRCHASHQOPLAARGVAARPGTLHDPFHSHEIAEQTLTLDALFLYIEHP 871  
QY 740 EYIDDLF-KLRKTSKANIKREBEVINOETFWVASEILRETNOLKMKIIKFIKIALHC 798  
Db 872 EYL--LMAKQNEKSPNLTQFTEHFENMNSYVWRSLIMLOEKQODERLLKIKIMKH 929  
QY 799 RECKNNSMFATISGLNLAVALRTTWELPKYKRLFDLODLDPSPRNMAKYNNVLN 858  
Db 930 RKLNNNSYLAIALSALDAPIRLLE--WQR---QTSGLAEYCTLIDSSSPFRAYRAALS 984  
QY 859 SQMLQPIIPLPFPVKKDLTFLEHGNDSKVDGLVNEKELMKIAKEIRHNGRMAVSMDPA 918  
Db 919 LMFRTKKKMSLSLSGSGTNATVLDVAQ--TGCHKRRVRRSSFLNAKKLYEDAQWARK 976  
QY 1019 -----SKRMQGE-----NILDSMRCFOGAHYETIRNDIIN----- 1049  
Db 977 VKOYLSNLELMDDESLOTLSLOCPEATNTLPKNPGDK 1015  
QY 1050 --FNFDFSDHLAEALWELSLIKIKRPNITRRKTDREK 1085

## RESULT 13

A38985  
nucleotide exchange factor CDC25 - human  
N:Alternate names: Ras-specific guanine nucleotide-releasing factor  
C:Species: Homo sapiens (man)  
C:Date: 19-Jul-1996 #sequence\_revision 19-Jul-1996 #text\_change 05-Nov-1999  
C:Accession: A38985; A46199; I58371  
R.Wei, W.; Broek, D.  
submitted to Genbank, December 1994  
A:Description: Cloning and analysis of the full length human cdc25 cDNA, a ras-specif  
A:Reference number: A38985  
A:Accession: A38985  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1275 <WEI>  
A:Cross-references: GB:I26584; NID:9433719; PIDN:AA58417.1; PID:9433720  
R.Wei, W.; Mosteller, R.D.; Sanyal, P.; Gonzales, E.; McKinney, D.; Dasgupta, C.; Li,  
Proc. Natl. Acad. Sci. U.S.A. 89, 7100-7104, 1992  
A:Title: Identification of a mammalian gene structurally and functionally related to  
A:Reference number: A46199; MUID:92357779; PMID:1379731  
A:Accession: A46199  
A:Molecule type: mRNA  
A:Residues: 1047-1054,'A',1056-1112,'G',1114-1135,'C',1137-1275 <WE2>  
A:Experimental source: brain  
A:Note: sequence extracted from NCBI backbone (NCBIN:111098, NCBIP:111099)  
R.Schneidhoffer, F.; Faure, M.; Fath, I.; Chevallier-Multon, M.C.; Apitou, F.; Dutrill  
Oncogene 8, 1477-1485, 1993  
A:Title: Identification of a human guanine nucleotide-releasing factor (H-GRF5) spec  
A:Reference number: I58371; MUID:93275641; PMID:7684828  
A:Accession: I58371  
A>Status: preliminary; translated from GB/EMBL/DDJ  
A:Molecule type: mRNA  
A:Residues: 787-1275 <RES>  
A:Cross-references: GB:S62035; NID:9386046; PIDN:AA826881.1; PID:9386047  
C:Superfamily: CDC25-type guanine nucleotide exchange activator homology; CDC24 homol  
F:240-426/Domain: CDC24 homology <CD24>  
F:1036-1272/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 3.8%; Score 290; DB 2; Length 1275;  
Best Local Similarity 19.4%; Pred. No. 9,7e-07;  
Matches 218; Conservative 162; Mismatches 393; Indels 348; Gaps 48;  
QY 3 PLATPANHCWGGO-----EKHSLPADFTKL-HLTDSLHQYVHVSSH-----S 46  
Db 386 PRYITLHDVLAHPHEHVERNSLDYAKSKLELSRIMHDEVSETENIRRNLAIERMIE 445  
QY 47 GCSITSDSGSSSL--SDIYATSEAGDMDSLGLPETAVSEDDDDDEDIERASDPLMSR 104

```

Db 446 GCETLLDTSQTFVROGSLIQVPMSEKGIITRGLSLSLEKES-----ERQCF-LFSK 497
QY 105 DIV-----RDLKEDPIDRTD-----DIEQL-----LEFMQ 132
Db 498 HLITCRSGSGKLHLTKNGVISLIDCTLLEEPESTEBAKSGODIDHDFKIGVEPKDS 557
QY 133 LPAFAMNTMSVRELCVAVMFVAVERAGTIVLNDGFELOSMVSLGVYVTPDCKAEI 192
Db 558 P-----FTVI-----LVASSROKAMWISDISQCV-----NI 586
QY 193 LCMG-----NSFGVPTMDKEVMKGMVARTV--DDCOFVCIADQYCRILNOYEKNM 242
Db 587 RCNGLMNNAEENSKYIVPQMIKRTREGTEAEKMSRSDASLYC-----DDVIDRF 636
QY 243 QKVEEGEIYVAKHEHRLDRTGTRKGITVKGISERLTMHVEHSVVDPTFIEDLLTY 302
Db 637 SKTMNSCKVQIIR-----YASVERLERLT-----DLRFLSIDFTLNTFLHST 678
QY 303 RTFLSPMEYGGKLEWFENDPSLRDKVTRVLLMV-----NNHFNDEGDPAMT 351
Db 679 RVF-TTAVVLDKLITTYKKP-----ISALPARKLRLSLELLFASGQNNKL--LYGEPPKS 730
QY 352 -RLEEFEN--NLEREKMGHLRLNLINACAKARRMLTLTKPSREAPLFFILGG---- 404
Db 731 PRATRKFSPPPLSITKTSSPSR-----RRKLISINIP-----IITGKALD 772
QY 405 -----SEKGFIFVSDVSGSKATEAGLKGQIILEVNGQNFENIQLSKAMEILANNT 457
Db 773 LAGSLSCNSNGYTSYMSAMSPFSKAT-----LDTSK----- 803
QY 458 HLTSTVNTLNFVKELLRLSEKRNCAPIHLPIKIDI--KKASHYSTPDLAVDEQYIGL 515
Db 804 -----LYSSSFFNKIPDE--GDPTPEKEDPSALSKOSSSVMSREES--DIDQ---- 848
QY 516 EKVKKSKANTVGGRNK--LKTILDKTRISIDPKPYNDIGIOSODDSVIGLRQTKHI 572
Db 849 ---NQSDGDGTETSPTKSPPTPKGVKNKNSSEPLFESYNN--GYVMTSCRELNNRSALSA 904
QY 573 PTALPVSGTSSSNPDLQSHHRLDFSATPDLDPQYLRFVKADQOS--RYIMISKOTTA 630
Db 905 ASAALINAGNEGTPNKEKTRMSLASAG-----FPPROKNGDKFVIRRAATN 954
QY 631 KEVYIOAIRFAVATATPD---QYSLCEVYTPREGVIKORRLPDOLSKLADRIQLSGRY 686
Db 955 R--VLNVLRRHVNSKHSODEFTNDELKCKVIGFLBEVMDHPELLTOERKAANI----- 1005
QY 687 LKNMNETTLCSDDEDAQELRESQISLQL-----STYEVATQUSMNEF 731
Db 1006 -----IRTLQOEDPG-----DNOITLEEITQOMAGVKAEPFENHSALTEIAEQLTLDHL 1054
QY 732 LFRNIEPTETI-DLDFKL-----RSKTSQCANLREFEVIINOETFWVASEILRETQOLKRM 786
Db 1055 VFKKIPEFEFGQGMKLEKMERITYIMKTTHFNDSN-----LIASIIINEDIINARVS 1110
QY 787 TIKHEFIKIALHCRECKNENSMFAITISGLNLAVARLRTTWKLEPKNYEKLFOQIODELP 846
Db 1111 ALEKVVAVADICRCLHNNAVALEITSSMNRSAIFPLKTKTWKSVSQTALDLDKLOKLVS 1170
QY 847 SRNMAKYNNVLNSQLOPPIPLPFYIKKDLFLHEGDSKV--DGLVFEKLMIAKIER 905
Db 1171 EGRFNLEAL--KNCDDPCVPYLGMYLTDLAFIEEGPNTEDGLVNFMSKRMISHTIR 1228
QY 906 HVGRAVSNMPPALMFRTRRKKKRSLSLSOGSTNATVLDVAOTGHHKRYRRSSFLNAK 965
Db 1229 EI-----RQOQOATV 1238
QY 966 KLYEDAQARKVQOYLSNLELEMEDESLQTLSDCEPATNT 1006
Db 1239 KIEHOA----KVTOYLLDOSFVMEDESLYESLRIEPKLPT 1275

```

```

guanine-nucleotide-releasing protein - rat
N:Alternate names: CDC25 protein homolog
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999
C:Accession: S29083
C:Shou, C.; Farnsworth, C.L.; Neel, B.G.; Feig, L.A.
Nature 358, 351-354, 1992
A:Title: Molecular cloning of cDNAs encoding a guanine-nucleotide-releasing factor fo
A:Reference number: S29083; MUID:92250260; PMID:1379346
A:Accession: S29083
A:Molecule type: mRNA
A:Residues: 1-1244 <SHO>
A:Cross-references: EMBL:X67241; NID:957664; PIDN:CAA47666.1; PID:957665
A:Note: The authors translated the codon GAG for residues 135 and 137 as Glu
C:Superfamily: CDC25-type guanine nucleotide exchange activator homology; CDC24 homol
F:240-426/Domain: CDC24 homology <CD24>
F:1005-1241/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 3.7%; Score 288; DB 2; Length 1244;
Best local similarity 18.4%; Pred. No. 1.2e-06;
Matches 201; Conservative 171; Mismatches 368; Indels 350; Gaps 42;

QY 17 EKHSLPADPTKL-HLTDLSHPQVTHYSSSH-----SGCSITSDSGSSSL--SDIY 63
Db 405 ERNSLDYAKRSKLELSRYVHDEVESENIKRLALERMITEGCEILLDTSQTFVROGSLI 464
QY 64 QATESEAGMDLISGLPETAIVEDSD-----DDEEDIERASDPLMSRDIVR 108
Db 465 QVMSSEKGIKINKRGLSLSLAKKEGROCFLFKHLICTRGSGSKHLTKNGVIS--LI 521
QY 109 DC-LEKPIDRTDDIEQLIEFM-----HOLPAFAMNTMSVRELCVAVMFVAVF-- 157
Db 522 DCTLDDPEHMDGDGQGEVDHDLFKIWEPEKSPPTVILVASSQOEKAAMTSDIQC 581
QY 158 ---RAGIYVNDDEELDSVSLVNGSVETYPD---GKAELICMGNSFGVSPMDKEYMK 211
Db 582 DNIRCMGLMNAEE-----NSKYIVPQIMSDSLSCDDVDYDFSKTMNS----- 627
QY 212 GVMKTKVDDCFQVCIADQYCRILNOVEKNMKQVEEGEIVAKHEHRLDRTGTRKGITV 271
Db 628 -----CKVL-QIR-----YAS 637
QY 272 IKTSERLMHVLEHSVNDPTFIEDFLTYRTFLSPMEVGKLEWENDPSLRKIVR 331
Db 638 VERLERLT-----DLRFLSIDFLNFTLHSTRVF--DVAVVLDKLSTYKRPITAIAPAS 691
QY 332 VLLMNVNH-----FNFEGDPAMTRFLEFENNLEREKMGHLRLNLINACAKARRML 386
Db 692 LELLFSSSHNTKLLYGAPKSPRASR--KFSSPP-----LAIGTSSPVARRKL 738
QY 387 TLTKPSREAPLFFILGSGSEKGFIEVDVSDGSKATEAGLKGQDILEVNGQNFENIQL 446
Db 739 SLNIP-----IITGG----- 748
QY 447 SKAMEILR-----NNTHLST-----TVKTNLFVKELLRLRSEKRNCAPIHLPIK 492
Db 749 -KAEELASLSCPSDGYNNHSPISFPGKTLTDTSKLCVASSLRTREE-----IDMT 799
QY 493 DIKASRYSPDLAVDEQYVIGLEKVNKSKANTVGGRNKLKRLDKTRISLIPQRPYND 552
Db 800 TLRESSGFRKP-----TSDLIKESDDQSD-----VDDTEVSPPTPKSFRN 841
QY 553 IGIGOSODDSIVGLRQTKHPIPTALPVSGTSSSNPDLQSHHRLDFS----- 601
Db 842 -----RITQEPFLFENYNGSIMTGR--DLMDSNRSPLSATSFAIATAGANE 886
QY 602 TPDLPDQVLRV-----FKADQOS--RYIMISKOTTAKEVYIOAIRFAVATATPD----- 648
Db 887 SPANKEIYRRMSLANGTGSSDQNRIDKEFYIRATNR--VLNVLRRHWTKHSODEFTD 944
QY 649 --OYS-LC-----EVSVTPREGVIKORRLPDOLSKLADRIQLSGRYLYKNNMETELCSDEDA 702
Db 945 LKTKVIGCFEEVYVHNDPLRQERKKAANIMRTLTQREIENH-----SMDELMLMTEGV 1000

```

F:65-123/Domain: SH3 homology <SH3>  
F:1301-1542/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 3.7%; Score 282.5; DB 1; Length 1589;  
Best Local Similarity 22.1%; Pred. No. 3.3e-06;  
Matches 133; Conservative 108; Mismatches 244; Indels 117; Gaps 22;

```
QY 333 VILWVNNHNFNDPEDPATREFL-----EPEENNLEEKKGHGHLRLINAC----- 377
Db 988 VLVSDVDFEENNLE-----FLNNSDPKDKCEK-----YVEISNLACIIYVDLIEREN 1035
QY 378 ----AAKARRLMTLTKPSREAPLPFILLGSGSEGFIF-VDSVDSGSKATEAGLRKGDQ 432
Db 1036 LLNVAARMKNNLTAE-----LLKGQGEKMFDIYSDYSDDDSENDEAI--DDE 1083
QY 433 ILEVNGQFENIQLSKAMEILRNNTHTUSTYVKTMLFVFKELLTLPSEKRNAGAPHLPIG 492
Db 1084 L-----GSEDIYIERKKAIMEKN-----LPWFILSDYETSLVYDSGKIRGTKEA 1128
QY 493 DKKARSYIPDLAVDYOVIQGLEKVNKKSKANTVGGKNNLKKILDKTRISILIPQK--PY 550
Db 1129 LIEHTSELVDAAFNVTMLTFRSI-----LTTREFFYAI--YRINLYPPEGLSY 1178
QY 551 NDIGIGOSODSIYGLRQTKHIPTALPVSGTLSSSNPDLLQSHRILDFSATPDLPQVL 610
Db 1179 DDYNIWEIKSKNPICRQVNIIMRTFLQYWRNRYEPIG----PLINPAK----- 1225
QY 611 RVFKADQOSRIIMISKDTTAAEVYIQAIRPAVATATPDQVSLCEVSATPEGVTKQRRLPD 670
Db 1226 -----MVYSEKIIGAEDLLQKINEKLINEKE-----PYDP---KQO---D 1261
QY 671 QLSKLADRIQLSGRYLLKNNMETETLCSDEDAQELRESQISILQSTVEVATQLSMRNF 730
Db 1262 SVSAVVOYTKRDNKSPIL--HMSSSLPSGA--SSAFRLKKIKLLDIDPYATQGLVLEH 1318
QY 731 ELFRNIETETIIDL--KLRSKTSKANLKRFEVINOETFWVAASEILLRETNOLKRMKIT 788
Db 1319 DLYRITMFECLIDRAMGTRKCYCMGSGSPNITKFINANMTLTFVSHITIVKQADVKTRESKLT 1378
QY 789 KHFIKIALHCRCKNFENMFALISGLNLAPARLRTTWEEKLPNXYEKLFDLODLFPDSR 848
Db 1379 QYFVIVAOHCKELNENFSSMTAIVSALYSPIYRLKRTMDVSTESKDLKLNLNLMDSK 1438
QY 849 NNAKTRNVLNSONLOPILPLFVPIKKDLTFLEHGNDG--KYDGLVNFEXLRMIAREIR 905
Db 1439 NFVKYRELLRSYT--DVACVPFEGVYLDLTFYVGNDFLHNSINIINFESKRTKIANIVE 1497
QY 906 HV 907
Db 1498 EI 1499
```

Search completed: February 18, 2003, 21:27:08  
Job time : 72 secs

**BLANK PAGE**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: February 18, 2003, 21:04:11 ; Search time 95 seconds  
(without alignments)  
3251.207 Million cell updates/sec

Title: US-09-911-826A-2

Perfect score: 7721

Sequence: 1 MKPLAIPANHCVMGQCEKHS.....PYQSQGFSTEDEDEQVSAAV 1499

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	7721	100.0	1499	4	O9Y4G8
2	4626.5	59.9	1601	4	O8TEU7
3	4473.5	57.9	1509	4	O96PC1
4	4060	52.6	1391	4	O8TEU6
5	3871.5	50.1	1204	4	O9UHV4
6	3572.5	46.3	1113	4	O8TEA3
7	2358.5	30.5	834	11	O8R3E5
8	2352.5	30.5	1551	5	O9VMF3
9	2352.5	30.5	1573	5	O95V18
10	2173	28.1	579	4	O43386
11	1798.5	23.3	1347	5	O95WR8
12	1798.5	23.3	1470	5	O21218
13	1779	23.0	1311	5	O95NL8
14	1021.5	13.2	528	11	O8VCX0
15	677	8.8	876	11	O8VCC8
16	676	8.8	881	4	O95398

17	676	8.8	884	11	O92IC8	O92IC8	rattus norv
18	675	8.7	881	4	O8WVNO	O8WVNO	homo sapien
19	671	8.7	881	4	O95634	O95634	homo sapien
20	666.5	8.6	993	11	O92IP0	O92IP0	mus musc
21	666.5	8.6	1011	11	O9E026	O9E026	mus musc
22	665.5	8.6	1011	4	O8WZ22	O8WZ22	homo sapien
23	664.5	8.6	1011	4	O95636	O95636	homo sapien
24	659.5	8.5	1011	4	O8TAA4	O8TAA4	homo sapien
25	658.5	8.5	696	11	O8VIP9	O8VIP9	mus musc
26	654.5	8.5	699	11	O9CW52	O9CW52	mus musc
27	649	8.4	861	5	O95U79	O95U79	drosophila
28	637	8.3	957	5	O9V9A3	O9V9A3	drosophila
29	455	5.9	96	4	O9UFT5	O9UFT5	homo sapien
30	447.5	5.8	580	4	O92565	O92565	homo sapien
31	433	5.6	436	11	O92IC7	O92IC7	rattus norv
32	431	5.6	456	4	O9UHV5	O9UHV5	homo sapien
33	392	5.1	287	11	O8R1R1	O8R1R1	mus musc
34	365	4.7	1721	5	O8SSQ0	O8SSQ0	dictyosteli
35	354.5	4.6	1592	5	O8SSU1	O8SSU1	dictyosteli
36	354.5	4.6	1596	5	O969D6	O969D6	drosophila
37	327	4.2	1575	5	O8SSN3	O8SSN3	dictyosteli
38	321.5	4.2	929	5	O8SSQ3	O8SSQ3	dictyosteli
39	321.5	4.2	1048	3	O12037	O12037	saccharomyc
40	299.5	3.9	426	5	O8P9J6	O8P9J6	drosophila
41	299.5	3.9	612	5	O9VNR5	O9VNR5	drosophila
42	298.5	3.9	1359	5	O8T6G6	O8T6G6	dictyosteli
43	295	3.8	1086	11	O91Z22	O91Z22	mus musc
44	294	3.8	1077	4	O13905	O13905	homo sapien
45	292	3.8	1413	5	O9NBD3	O9NBD3	caenorhabdi

## ALIGNMENTS

RESULT 1  
ID O9Y4G8 PRELIMINARY; PRT; 1499 AA.  
AC O9Y4G8;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE KIAA0313 protein.  
GN KIAA0313.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRN;  
RX MEDLINE-97349984; PubMed-9205841;  
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,  
RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. VII.  
RT The complete sequences of 100 new cDNA clones from brain which can  
RT code for large proteins in vitro.";  
RL DNA Res. 4:141-150(1997).  
DR EMBL; AB02311; BAA20772.1; -  
DR InterPro; IPR000595; CNMP\_binding.  
DR InterPro; IPR001478; PDZ.  
DR InterPro; IPR000651; RasGEFN.  
DR InterPro; IPR001895; RasGEFN\_CDC25.  
DR InterPro; IPR000159; RA\_domain.  
DR Pfam; PF00027; CNMP\_binding; 1.  
DR Pfam; PF00595; PDZ; 1.  
DR Pfam; PF00788; RA; 1.  
DR Pfam; PF00617; RasGEFN; 1.  
DR Pfam; PF00618; RasGEFN; 1.  
DR SMART; SM00100; CNMP; 1.  
DR SMART; SM00228; PDZ; 1.  
DR SMART; SM00314; RA; 1.  
DR SMART; SM00147; RasGEFN; 1.  
DR SMART; SM00229; RasGEFN; 1.

DR PROSITE: P550042; CNMP\_BINDING\_3; 1.  
 DR PROSITE: P550106; PDZ; 1.  
 SQ SEQUENCE 1499 AA; 167416 MW; 1909E8A12637E001 CRC64;

Query Match 100.0%; Score 7721; DB 4; Length 1499;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1499; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKPLAIPANHGVMGQOEKHSIPADFTKLHLTDSLHPQVTHVSSHSGCCSTSDSGSSSL 60  
 1 MKPLAIPANHGVMGQOEKHSIPADFTKLHLTDSLHPQVTHVSSHSGCCSTSDSGSSSL 60  
 DB 1 MKPLAIPANHGVMGQOEKHSIPADFTKLHLTDSLHPQVTHVSSHSGCCSTSDSGSSSL 60

QY 61 DIYQATESEAGMDLSGLPTAVDSDEDDDEEDIERASDPLMRDLYRDCLEDPIDRTD 120  
 61 DIYQATESEAGMDLSGLPTAVDSDEDDDEEDIERASDPLMRDLYRDCLEDPIDRTD 120  
 DB 61 DIYQATESEAGMDLSGLPTAVDSDEDDDEEDIERASDPLMRDLYRDCLEDPIDRTD 120

QY 121 DDIEQLLEFPHQOLPAFANMMSVRELCAMVFAVERACTIVLNDGEELDSVILNGS 180  
 121 DDIEQLLEFPHQOLPAFANMMSVRELCAMVFAVERACTIVLNDGEELDSVILNGS 180  
 DB 121 DDIEQLLEFPHQOLPAFANMMSVRELCAMVFAVERACTIVLNDGEELDSVILNGS 180

QY 181 VEVYTPDGKAEILCMGNSFGVSPPTMDKEYMKGVWRTKVDCCQFVCI AQODYCRILNQVEK 240  
 181 VEVYTPDGKAEILCMGNSFGVSPPTMDKEYMKGVWRTKVDCCQFVCI AQODYCRILNQVEK 240  
 DB 181 VEVYTPDGKAEILCMGNSFGVSPPTMDKEYMKGVWRTKVDCCQFVCI AQODYCRILNQVEK 240

QY 241 NMQKVEEGEIVMYKHEIRELDRGTGRKGHIVIGTSERLTMHLVEHSVVDPTFIEDFLL 300  
 241 NMQKVEEGEIVMYKHEIRELDRGTGRKGHIVIGTSERLTMHLVEHSVVDPTFIEDFLL 300  
 DB 241 NMQKVEEGEIVMYKHEIRELDRGTGRKGHIVIGTSERLTMHLVEHSVVDPTFIEDFLL 300

QY 301 TYRFTLSPPEVGVKRLLEWNPDSLARDKVT RVVLLWVNNHNFDEEGDPAMTRLEEEENN 360  
 301 TYRFTLSPPEVGVKRLLEWNPDSLARDKVT RVVLLWVNNHNFDEEGDPAMTRLEEEENN 360  
 DB 301 TYRFTLSPPEVGVKRLLEWNPDSLARDKVT RVVLLWVNNHNFDEEGDPAMTRLEEEENN 360

QY 361 LERKMGHGLRLNLIACAAKAKRRLMTLTPRSRAPLPFTLLGSEKGFGLPVDVSUGS 420  
 361 LERKMGHGLRLNLIACAAKAKRRLMTLTPRSRAPLPFTLLGSEKGFGLPVDVSUGS 420  
 DB 361 LERKMGHGLRLNLIACAAKAKRRLMTLTPRSRAPLPFTLLGSEKGFGLPVDVSUGS 420

QY 421 KATFAGLKRGDQILEVNGQNPENITOLSKAMEILNNHTLSTVNTNLFVKELTRLSEE 480  
 421 KATFAGLKRGDQILEVNGQNPENITOLSKAMEILNNHTLSTVNTNLFVKELTRLSEE 480  
 DB 421 KATFAGLKRGDQILEVNGQNPENITOLSKAMEILNNHTLSTVNTNLFVKELTRLSEE 480

QY 481 KRNGAPHLPKIGDIKKASRYSIDPLAVDVQVIGLEKVNKSKRANTVGGRNKLKILDKT 540  
 481 KRNGAPHLPKIGDIKKASRYSIDPLAVDVQVIGLEKVNKSKRANTVGGRNKLKILDKT 540  
 DB 481 KRNGAPHLPKIGDIKKASRYSIDPLAVDVQVIGLEKVNKSKRANTVGGRNKLKILDKT 540

QY 541 RISTLPQKPYNDIGIGSQSDSYGLQTKHPIPALPVSGTLSSNDPLQSHRRILDFS 600  
 541 RISTLPQKPYNDIGIGSQSDSYGLQTKHPIPALPVSGTLSSNDPLQSHRRILDFS 600  
 DB 541 RISTLPQKPYNDIGIGSQSDSYGLQTKHPIPALPVSGTLSSNDPLQSHRRILDFS 600

QY 601 ATPPLPOVLVFRKADOOSRYIMISKDTAKEVVIQAIRFAVATPDPQSLCEVSYTPE 660  
 601 ATPPLPOVLVFRKADOOSRYIMISKDTAKEVVIQAIRFAVATPDPQSLCEVSYTPE 660  
 DB 601 ATPPLPOVLVFRKADOOSRYIMISKDTAKEVVIQAIRFAVATPDPQSLCEVSYTPE 660

QY 661 GVIKORRLPDQLSKLADRIQLSGRYVLLKNNMETETLCSDEDAQELRESQISLLQSTVE 720  
 661 GVIKORRLPDQLSKLADRIQLSGRYVLLKNNMETETLCSDEDAQELRESQISLLQSTVE 720  
 DB 661 GVIKORRLPDQLSKLADRIQLSGRYVLLKNNMETETLCSDEDAQELRESQISLLQSTVE 720

QY 721 VATQLSRNLEFLFNIEPTETIDYDLFKLRSTSCANLKRFEVINQETFWVASILLRETN 780  
 721 VATQLSRNLEFLFNIEPTETIDYDLFKLRSTSCANLKRFEVINQETFWVASILLRETN 780  
 DB 721 VATQLSRNLEFLFNIEPTETIDYDLFKLRSTSCANLKRFEVINQETFWVASILLRETN 780

QY 781 QLKRMKIKHFKIKIALCHRECKNFSMFAISGLNLPAVALRTTWELPKRYEKLFDL 840  
 781 QLKRMKIKHFKIKIALCHRECKNFSMFAISGLNLPAVALRTTWELPKRYEKLFDL 840  
 DB 781 QLKRMKIKHFKIKIALCHRECKNFSMFAISGLNLPAVALRTTWELPKRYEKLFDL 840

QY 841 QDLFDPSSNMAYKNVNLSONLQPPILPFPVIKKDLTFLEHGNDSKVDGLVNEEKLMI 900  
 841 QDLFDPSSNMAYKNVNLSONLQPPILPFPVIKKDLTFLEHGNDSKVDGLVNEEKLMI 900  
 DB 841 QDLFDPSSNMAYKNVNLSONLQPPILPFPVIKKDLTFLEHGNDSKVDGLVNEEKLMI 900

QY 901 AKELRHVGRMASVNMVDPALMRTKRRKRSISLSSQSTNTATVDVQGTGKKRVRRSS 960  
 901 AKELRHVGRMASVNMVDPALMRTKRRKRSISLSSQSTNTATVDVQGTGKKRVRRSS 960  
 DB 901 AKELRHVGRMASVNMVDPALMRTKRRKRSISLSSQSTNTATVDVQGTGKKRVRRSS 960

QY 961 FLNAKKLYEDAQMARKVQYLSNLELMEDESLQTLSDCEPATNTLFPKNPDGKRPVNSE 1020  
 961 FLNAKKLYEDAQMARKVQYLSNLELMEDESLQTLSDCEPATNTLFPKNPDGKRPVNSE 1020

DB 961 FLNAKKLYEDAQMARKVQYLSNLELMEDESLQTLSDCEPATNTLFPKNPDGKRPVNSE 1020  
 1021 TSPVAPRAGSQOKAASLPQPOQPPPAHKINQGLVPAVSLYPSKKRYKVDLPFGGINS 1080  
 1021 TSPVAPRAGSQOKAASLPQPOQPPPAHKINQGLVPAVSLYPSKKRYKVDLPFGGINS 1080  
 DB 1021 TSPVAPRAGSQOKAASLPQPOQPPPAHKINQGLVPAVSLYPSKKRYKVDLPFGGINS 1080

QY 1081 PQALKITLSEEGSLERKKAQEDTISNASSOLSPPTSPQSSPRKGYTLAPSGTVDNF 1140  
 1081 PQALKITLSEEGSLERKKAQEDTISNASSOLSPPTSPQSSPRKGYTLAPSGTVDNF 1140  
 DB 1081 PQALKITLSEEGSLERKKAQEDTISNASSOLSPPTSPQSSPRKGYTLAPSGTVDNF 1140

QY 1141 SDGHSIEISSRSIYNSNFFSDVPYSLHDERORHSVSIVETNLGMRMERTMIEPDQY 1200  
 1141 SDGHSIEISSRSIYNSNFFSDVPYSLHDERORHSVSIVETNLGMRMERTMIEPDQY 1200  
 DB 1141 SDGHSIEISSRSIYNSNFFSDVPYSLHDERORHSVSIVETNLGMRMERTMIEPDQY 1200

QY 1201 SLGSYAPMSEGRGLYATAVISSPSTEELSDQGDRAVSLDAADSGRGSWTSCSSGSHDI 1260  
 1201 SLGSYAPMSEGRGLYATAVISSPSTEELSDQGDRAVSLDAADSGRGSWTSCSSGSHDI 1260  
 DB 1201 SLGSYAPMSEGRGLYATAVISSPSTEELSDQGDRAVSLDAADSGRGSWTSCSSGSHDI 1260

QY 1261 QTIQHRSWETLPEGHTHFDYSGDPAGLWASSHMDQIMFSDHSTKYRNQONRESLEQA 1320  
 1261 QTIQHRSWETLPEGHTHFDYSGDPAGLWASSHMDQIMFSDHSTKYRNQONRESLEQA 1320  
 DB 1261 QTIQHRSWETLPEGHTHFDYSGDPAGLWASSHMDQIMFSDHSTKYRNQONRESLEQA 1320

QY 1321 QSRASWASSTGWGEDSEGDTGTRKRGKDYSEIASSSLTSVTTEETKVPMPAHIAV 1380  
 1321 QSRASWASSTGWGEDSEGDTGTRKRGKDYSEIASSSLTSVTTEETKVPMPAHIAV 1380  
 DB 1321 QSRASWASSTGWGEDSEGDTGTRKRGKDYSEIASSSLTSVTTEETKVPMPAHIAV 1380

QY 1381 ASSTYGLARKKRGYREPPPPGYIGITIDFPGHSHPARPPDYVALQSRMVAR 1440  
 1381 ASSTYGLARKKRGYREPPPPGYIGITIDFPGHSHPARPPDYVALQSRMVAR 1440  
 DB 1381 ASSTYGLARKKRGYREPPPPGYIGITIDFPGHSHPARPPDYVALQSRMVAR 1440

QY 1441 SSPTAGSSVQOPHGHPTSSRPVKNPOMHKPNESDRLAPYOSGCFSTEDEDEQYSAV 1499  
 1441 SSPTAGSSVQOPHGHPTSSRPVKNPOMHKPNESDRLAPYOSGCFSTEDEDEQYSAV 1499  
 DB 1441 SSPTAGSSVQOPHGHPTSSRPVKNPOMHKPNESDRLAPYOSGCFSTEDEDEQYSAV 1499

RESULT 2  
 08TEU7 PRELIMINARY: PRT: 1601 AA.

AC 08TEU7;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Rap1 guanine nucleotide-exchange factor PDZ-GEF2A.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kuiperij H.B., de Rooij J., Reihmann H., van Triest M.,  
 RA Wittinghofer A., Bos J.L., Zwartkruis F.J.T.;  
 RT "Characterization of the PDZ-GEFs, a family of guanine nucleotide  
 exchange factors specific for Rap1 and Rap2."  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: AF478468; AAL79915.1; -  
 SQ SEQUENCE 1601 AA; 179406 MW; E69D208431909A0E CRC64;

Query Match 59.9%; Score 4626.5; DB 4; Length 1601;  
 Best Local Similarity 64.1%; Pred. No. 1,4e-258;  
 Matches 959; Conservative 163; Mismatches 299; Indels 75; Gaps 26;

QY 20 SLPADFTKLHLTDSLHPQVTHVSSHSGCCSTSDSGSSLSDIYQATESEAGMDLSGLP 79  
 165 SLPADFTKLHLTDSLHPQVTHVSSHSGCCSTSDSGSSLSDIYQATESEAGMDLSGLP 224  
 DB 165 SLPADFTKLHLTDSLHPQVTHVSSHSGCCSTSDSGSSLSDIYQATESEAGMDLSGLP 224

QY 80 ETAVDSDDDEE-EDIERASDPLMRDLYRDCLEKDPIDRTDDIDOLEFPHQOLPAFAN 138  
 138 ETAVDSDDDEE-EDIERASDPLMRDLYRDCLEKDPIDRTDDIDOLEFPHQOLPAFAN 138  
 DB 80 ETAVDSDDDEE-EDIERASDPLMRDLYRDCLEKDPIDRTDDIDOLEFPHQOLPAFAN 138

QY 225 EGPVDSDEDEDEIDR-TDPLQGRDLVRECLEKEPADTDDIDOLEFPHQOLPAFAN 283  
 225 EGPVDSDEDEDEIDR-TDPLQGRDLVRECLEKEPADTDDIDOLEFPHQOLPAFAN 283  
 DB 225 EGPVDSDEDEDEIDR-TDPLQGRDLVRECLEKEPADTDDIDOLEFPHQOLPAFAN 283

QY 139 MTMSVRELCAMVFAVERAGTIVLNDGEELDSVILNGSVEVYTPDGKAEILCMGNS 198  
 139 MTMSVRELCAMVFAVERAGTIVLNDGEELDSVILNGSVEVYTPDGKAEILCMGNS 198  
 DB 139 MTMSVRELCAMVFAVERAGTIVLNDGEELDSVILNGSVEVYTPDGKAEILCMGNS 198

QY 284 MTMSVRELCAMVFAVERAGTIVLNDGEELDSVILNGSVEVYTPDGKAEILCMGNS 343  
 284 MTMSVRELCAMVFAVERAGTIVLNDGEELDSVILNGSVEVYTPDGKAEILCMGNS 343  
 DB 284 MTMSVRELCAMVFAVERAGTIVLNDGEELDSVILNGSVEVYTPDGKAEILCMGNS 343









Qy	798	CRCKNFEMSAIIISGLNLAPVARLRITWTKKLPNKYEKLEFLODLQDLFDPSSRMATYRNVL	857
Db	991	CRECKNFEMSAIIISGLNLASVARKRTWKLEPSYEKLEHLLQDLQDLFDPSSRMATYRNVL	1050
Qy	858	NSQNLQPIITPLPFVIAKDLTLFHEGNDSSVQDGLNVEFKRLIAKEIRHVGMAVYNDDP	917
Db	1051	SSQSQMPIITPLPFVYVAKDMLFHEGNDSSVQDGLNVEFKRLIAKEIRHVGMAVYNDDP	1110
Qy	918	ALMRTTRKKKRRSLGSLSSQSTNATYVLDAQAQTGGHKKRVRRSSFTLNAAKLYEDAAQARKV	977
Db	1111	AMERQR-----SLSSQSTNSNMIDV-OGGAHKRRARRSSLLNAAKLYEDAAQARKV	1161
Qy	978	KQYLSNLELEWDEESLTLSELOEPTNTLTPKKPGGKRVK	1018
Db	1162	KQYLSNLELEWDEEKFQMSLQWEPAYGTLTKNLSEKRSK	1202

RESULT 6	
Q8TEA3	
ID	Q8TEA3
AC	Q8TEA3; PRELIMINARY; PRT; 1113 AA.
DT	01-JUN-2002 (TrEMBLrel. 21, Created)
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	CDNA FLJ33738 fis, clone HEP15081, highly similar to PDZ domain-containing guanine nucleotide exchange factor 1.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX	NCBI_Taxid=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.
RA	Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA	Tanaka T., Nakamura Y., Isozaki T., Sugano S.;
RT	"NEDO human cDNA sequencing project."
RL	Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
DR	EMBL: AK074318; BAB5049.1; -;
SQ	SEQUENCE 1113 AA; 126109 MW; B12B7DFDE08C502B CRC64;

Query Match	46.3%	Score 3572.5	DB 4	Length 1113
Best Local Similarity	70.1%	Pred No. 6.8e-198		
Matches 718	Conservative 104	Mismatches 159	Indels 43	Gaps 14

QY	141	MSVRELCAVWVFEVWAGTIVYINDSGELDSWVILNGSVYVYVPGKMEILCMGMSFG	20
Db	1	MSVRELCASWVIEFVWQAGAILLEDQOELDSWVYILNGVYVISHPGKMEILCMGMSFG	60
QY	201	VSPIMDKYKGMVIRKTVDDCQFVCAAOODYCRILNOVERKMKOVEEEGEIVWVKEHREL	260
Db	61	ITPLIDLOYMHGIVRTYVDDCQFVCAAOODYRILNHVEKTHKVEEGEIVWVHEHREL	120
QY	261	DRTGRKGHTVYKGTSERLTMHVLVEASVVDPTTIEDPFLITRYFLSSPMEVCKLLMEF	320
Db	121	DRTGRKGHTVYKATPERLTMHLLIEBSIVDPTTIEDPFLITRYFLSPMDVGTIKLMEF	180
QY	321	NDSLRDKVTVVLLVWNNHFNDEQBPAMTRLEEFENNLEREKMGHRLINIACAK	380
Db	181	KIDSLRDKVTVRYVLLVWNNHFNDEQBPAMTRLEEFENKLEDTKMGHRLINIACTL--	238
QY	381	AKRRLM---LITPSSRAPLPFLILGSEKGFIEYDSVDSGSKATFAGLRKQDILTEVN	437
Db	239	--ORLMTGCAAGFPFVPSQFSLNGSEKGFIEYGVGEVPSKADSGLRKQDIMEVN	296
QY	438	GQNFENITQSKAMEILRNHNLSTVYTNLFVREKELLTRLSEKRCNCAHPRLKIGIKKA	497
Db	297	GQNFENITTFPKAAEILRNHNLHLVYTNLFVREKELLFR--TEDEKSGVPHPIPIAE--KKS	354
QY	498	SRYSIDPLADVQVIGLEKVVNKKSKANT--GGNRKLTLLDTRISILQKRYNDIGIG	556
Db	355	NRRSIGHVEDIDGA--SQEKGSKKVVANYVSGGRNKRIRKLLDTRISILPRLKLFSDGGLS	413

QY 557 QOSDDSTVGRKRLKPTALPRTDSTSSNPDLLQSHHRLDTSAPRDLDDOYLVRKAD 616  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db 414 QOSDDSTVGRKRLKPTALPRTDSTSSNPDLLQSHHRLDTSAPRDLDDOYLVRKAD 473  
QY 617 QOSRYIMISKDITAKEVYVQAIREFAVTATPDQYSLCEVSVTEGVYIKORRLPDQSLKLA 676  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db 474 QOSCYIIISIDTQAKAEVYFAVHVEFGITGASDYSLCEVSVTEGVYIKORRLPDQFSKLA 533  
QY 677 DRQLDGRVYLKNNMFEYLCSDPDQELLREQOISILQSLYEVATQLSMRNPELFRT 736  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db 534 DRQLDGRVYLKNNMFEYLCSDPDQELLREQOISILQSLYEVATQLSMRNPELFRT 593  
QY 737 EPTEYIDDLFKLRSKTSNCANLKRFEVYINQETFWASSEILRETNOILKRMKIIKHFIKIAL 796  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db 594 EPTEYIDDLFKLRSKTSNCANLKRFEVYINQETFWASSEILRETNOILKRMKIIKHFIKIAL 653  
QY 797 HCRECKNFSMFAIISGLNLAPVRLRTTWELPKNYEKLFDLQDLDPFSRMARYNY 856  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db 654 HCRECKNFSMFAIISGLNLAPVRLRTTWELPKNYEKLFDLQDLDPFSRMARYNY 713  
QY 857 LNSRONLQRPPIPEFYVKKDPTFLHGNDSKVQGLVNFELKRIAKELRHVGSMASYND 916  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db 714 LNSQSNRPPIPEFYVKKDPTFLHGNDSKVQGLVNFELKRIAKELRHVGSMASYND 773  
QY 917 PALMFRTRKKKMWLSLQSLSGSSTNATVLDVAQTGKHKRRRSFNNAKLYEDADQMARK 976  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db 774 PALMFRTRKKKMWLSLQSLSGSSTNATVLDVAQTGKHKRRRSFNNAKLYEDADQMARK 832  
QY 977 VQOYLSNLEBNDESLQTLSDCEPATNTLPKNPDGKKPVK-SETSVPAPRAGSOQKAO 1035  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db 833 VQOYLSNLEBNDESLQTLSDCEPATNTLPKNPDGKKPVK-SETSVPAPRAGSOQKAO 892  
QY 1036 SLFPOPOQOQPPAHKINQGLQVPAVSLYPRKKKYPVKDLPRFGINSQALKIILSLEEGS 1095  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db 893 HLHQF-----HRSQYLVQPAVNLHPHPRKKGQTNQD-BALNTSLPQO--KVLGTTEIS 941  
QY 1096 LERHKKQADDTISNMSQSLSPPTSPQSPRKGYYTLAPGTVDFNSDGHSEISSRSYIV 1155  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db 942 ---GKHTEDTISVASSLSHSPASPQSGPHK-----VGSF-----ISHSSKI 982  
QY 1156 SSSS 1159  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db 983 SSSS 986  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
RESULT 7  
Q8R3E5  
ID Q8R3E5 PRELIMINARY: PRT: 834 AA.  
AC Q8R3E5:  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)  
DE Similar to rrp guanine nucleotide exchange factor (Fragment).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
RN NCBI\_TaxID=10090;  
[1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC025553; AAH25553.1; -  
FT NON\_TER 1  
SQ SEQUENCE 834 AA; 92545 MW; E699BEB22FE6385 CMC64;  
Query Match 30.5%; Score 2358.5; DB 11; Length 834;  
Best Local Similarity 59.0%; Pred. No. 6.8e-128;  
Matches 520; Conservative 94; Mismatches 205; Indels 63; Gaps 19;  
Db 5 TAKEVYVQAIREFAVTATPDQYSLCEVSVTEGVYIKORRLPDQSLKLA  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
5 TAKEVYVQAIREFAVTATPDQYSLCEVSVTEGVYIKORRLPDQSLKLA  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
5 TAKEVYVQAIREFAVTATPDQYSLCEVSVTEGVYIKORRLPDQSLKLA



D	b	61	Y-ASLAEHQLDLSGLVESVDS-----DEEDLAESMDSLTVADANRDCLENDPARSEED	115
Q	y	123	IEOLLEFMHOLPAFANMTMSVREELCAVMFAVERAGITIVLNDGEEDMSVILNGSE	182
D	b	116	VEVLEFTEOGLKAFNTTTLAVRALCSVMFAVADKAGTVMSDEEDLSMVSLLINGAVE	175
Q	y	183	VTYDGAELLGNGNSGVSPIMDKRYMGVMYTKVDDQFCFACIAOQYCRILNGVEKNM	242
D	b	176	IEAHNGSREELQMGDSFGILLPTMDKLYHNGVMYTKCDDCFVCITQTDYRLOHGEBENT	235
Q	y	243	OKVEEE-GEIYMAKEHEHEDLRTGT-----RKGHIVIKTSERTYHML	283
D	b	236	RHDEDEGPFVWMTLELRISIGACTBDSNGSGSATGASALNMKRGHVYIRGPRELLDOL	295
Q	y	284	VEEHSVVDPTFIEDFLLTYRTFLSSPMGYKKLLEWFN-----DPSLRDQTVRVLL	335
D	b	296	VEEHSMDPTPYVEDEFLLTHRIFFIQNPQVYTSKLHWFLEQYDAKHQOGLRDRVTRVYLL	355
Q	y	336	WVNNHFFDEGDPAMTRFLEEFEENNLEREKMGCHLRLINACAANAARMLTLPSPSEA	395
D	b	356	WVNNHFFDEGDPAMTRFLEEFEENNLEREKMGCHLRLINACAANAARMLTLPSPSEA	415
Q	y	396	PLPILLGGSP-----KGFGLFVDSVSGSATFAGLKRGOILEVNGOFENIQ	445
D	b	416	PLNFRIVGTELRGVAITGMAAVGITISHVEGSKAODVGLKRGQIHEVNGSLDHYT	475
Q	y	446	LSKAMELIRNNTLSTVKNLFPVEKELTRLSE-----KRNAPHLP	489
D	b	476	SKRALTELTGTHLSISVKSNTLLGFPEIMQALEHGGCTAGSGSISAGSSSPFVSVPRI	535
Q	y	490	KIGCIK-----AS-----KSIDPLAVD-----	508
D	b	536	CANDIAKLHGKSDSTDELSVSASNRAMHVVRLSSVMDLDPDPCAPOTPREVSGSGNMA	595
Q	y	509	---VEOYIGLEKV-----KSKSAN-----TYGG-----RNMKLKILDKTRISILPQK	548
D	b	596	SNEPOOL-LOSVMNSAKKSGCNSNSDDODTKGGMYLAPKRRLOKALAKANL-----	647
Q	y	549	PYNDIGIGQSDDSIVGL-----RQTKHPIPALPYSG-----TLSSSNDL	589
D	b	648	L-NKQNHGSSLNDSODTLNDPDKSLSAVSSCSTQSSINGCIVGSGGLRVQSQSNPDL	706
Q	y	590	-----LOSH-HR-----ILDBSAT-----PDPQOYLRVKA	615
D	b	707	TSLNYDGSADGANGGGRLOVNYLNMHHRPSAASLTITNSQTHSLPDPYRPHVILKYKA	766
Q	y	616	DOOSRYTIMSDTJAKVYNOVIAIREFAVYATTPROYLSLCVSYTPBSVYIKORLPOLSKL	675
D	b	767	DOTCKYVILIKETTAHVYVMTLQDEGFIHDPSSNFFLCEVSGDGMVAKRPLPOLONL	826
Q	y	676	ADRILOSGRAYLKNMMETFLCSDDEQALDRESQJLSLQTSVEVATQDLSRKNELFRN	735
D	b	827	AERTSFARYYLKNDSYTEPLVDELALVELRSNVNPHLHNAVELAIDULTIOTDRANRQ	886
Q	y	736	IEPTYIDDLFKLSKTSKANLRFEEVINOETFWVASELIRETNOLKRMKILKHFIIKA	795
D	b	887	IESTEYDELFELRSRYGVPMLSKFAELVAVREMFVYSEICAEHNIVRMRKIVQOFIKIA	946
Q	y	796	LHCECKNFSMFAITISGLMLAPYARLRTTMEKLPKYEKLEODLODLFDPBSRNNAKTRN	855
D	b	947	RHCCECKNFSMFAIVSGLGHGAVSRLQRMELPKSYORLEFNDLODLMDPSRNMSKYRQ	1006
Q	y	856	VLSNQN-L-OPPIILFPVVIKDLFLHEGDSVYDGVNPEKRLMYAKELIRIVGMAVSN	914
D	b	1007	LVSHELLAQHPIIFYIYVKKDLTFTHLNDTVYDGLINFEKRLMYAKVRYLLTHMCSPP	1066
Q	y	915	MD-----PALMERTRKKKWRSLGSLSGOSTNA-----TVLVAOATGCH-----KRVRRSSFLTA	964
D	b	1067	YDLISLIELKQSGSNMFLSNQMSASQSNAAAGTV--IAANAGQATIKRKRKSTJAAPNP	1124
Q	y	965	KKLYEDQAMARKQVQYLSNLELEMDSESLQTSLOCERA-----TJTLK-----	10095

Db	1125	KKMEEEAQNVRPAKAYLNSLKLITLDEDLHKFSLCECPAHGSTYSGSISHGNTSHRSGG	1184
Qy	1010	-----NGDKRPVKSEIS-----PVAPRAGSQOK	1033
Db	1185	GSISGAGGSGGGGGSSSLNAGDGLSTYSHTSSSAANSSLSLKRHPSPPLTSTSS	1244
Qy	1034	AQSLPQEQOQPPRAHKINQSLQPAVSLYPSRKRVYVKDLPEFGINSIPALKILSLSEE	1093
Db	1245	TSSTISDHQRR-----QHNNNG-----PKEGIASPQAVKMLLSLES	1286
Qy	1094	GSLEHRKQKQAEADTISNASSQLSS--PPTSQSSPRKGYTLAPSGTYDNFSQSGHSEISSR	1151
Db	1281	SKINPH-----QPFVPRHGSTMAGVTPRLHMHMAHGFSTPSPGVV-----TSPA	1326
Qy	1152	SSIYSNSESFDSVPVSLHDERQRHNSIYEETNIGMRMERPMIEPD-----QYSLGSA	1206
Db	1327	TSVAANVOCTPSP--SPCSHRRRLASGNNIIIPRAIHERSHSDPAPRPPPLPSYDLSES--	1383
Qy	1207	PMSGRCLYATATYVISPSPTEELSSQOQGRASLDADSDRGSWTSCSSGSHNIDTIOH	1266
Db	1384	-----SSVTT--FRDLPLKSYTS-----GSISCDSS-----YVHQ	1413
Qy	1267	RSWETLPFGHTHEPDYSGDPAGLWASSSHMDQIMFSDHTKYRNKONRSFSLDAQSRASM	1322
Db	1414	QO-----HYHLY-----QOQOQNSQSHESPVPY	1439
Qy	1327	ASSTGWCEDESGDTGTRKRGKDVSIENASSLTSVTTEETKVPMP---AHIAVAS	1382
Db	1440	TAA-----DCRLIQISNNAVTARN--NSPCQSTWPTPSTPPPPNPTATIQLSA	1488
Qy	1383	STTGILARKREGRR-----PPPTPPYIGIPITDPREG--HSHPRARKPPDINV	1430
Db	1489	PTPAAYMHAARSQIQQLQOQOQSILAMPPEPPPY-NVP-----PLGSIYSH-----	1533
Qy	1431	ALQSRMVAASSDTPAGSPSYQOPHGH	1456
Db	1534	-----HQTAGSRHLNHHNGH	1549
RESULT 9			
RP	095V18	PRELIMINARY;	PRT: 1573 AA.
RA	095V18		
RT	"dizzy, a novel Drosophila guanine nucleotide exchange factor."		
RL	Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.		
DR	EMBL: AF134686; AAL28130.1; -		
DR	FLYBase: FBgn0031798; dizzy.		
DR	InterPro: IPR000595; GMP_binding.		
DR	InterPro: IPR001478; PDZ.		
DR	InterPro: IPR000651; RaseGEFN.		
DR	InterPro: IPR001895; RaseGRF_CDC25.		
DR	InterPro: IPR000159; RA_domain.		
DR	Pfam: PF000027; GMP_binding; 1.		
DR	Pfam: PF00595; PDZ; 1.		
DR	Pfam: PF00788; RA; 1.		
DR	Pfam: PF00617; RaseGEF; 1.		
DR	Pfam: PF00618; RaseGFN; 1.		
DR	PROSITE: PSS0042; GMP_BINDING_3; 1.		
DR	PROSITE: PSS0106; PDZ; 1.		
QO	SEQUENCE 1573 AA; 171796 MW; 4448811E7621BBFIC CRC64;		

Query Match 30.5%; Score 2352.5; DB 5; Length 1573;  
Best Local Similarity 36.2%; Pred. No. 4.1e-127;  
Matches 621; Conservative 225; Mismatches 458; Indels 413; Gaps 53;

QY 3 PLAIANHGVMGQOEKHSPLADFTKLHLNDSLHPQVTHVSSHSGSITSDDGSSLSLDI 62  
DB PTIAGTGVVGGSSSTINRP-----ELHQKCNR--GSHSPDTSXAGSGSDTMASN 60  
QY 63 YQATESEAGMDLSGLPETAVIDSEDDDDERIRASDPLMSRIYVDCLEKPIDRTDD 122  
DB Y-ASSUEAEIDISGLVESVDS---DEEDLAEISMDSLTVRAVDRCKLEKDAERSEED 115  
QY 123 IEOLLEPMHOLPAFANMTSVRRELCAVMYFAVERAGTIVLDNGEELDSMTLNGSVE 182  
DB VELLLEFTQGLKFTNITLAVRALCSVMYFAVADKAGTYVMSDGERLDSMTLNGAVE 175  
QY 183 VYTPDKAEILCKGNSPGVSPMDKREYMGVMRTKVDCCQVFCIAOQYCRLLNOYEKNM 242  
DB IEHANSRRELQMGDSFGILPTMDKLYHNGVMRTKCDCCQVFCITQDYRYRIHQGBENT 235  
QY 243 QXVEEE-GEIVMYKHEHREDTRGT-----RKGHVYKGTSERLTMH 283  
DB RREHDEGEVYVWTELRSLIGVGTDGSGSATGASASLNMKRGHVJLRGTEPERLQOL 295  
QY 284 VEEHSVVDPTFIEDFLTYRTFLSPMEVGKLLMEFN-----DPSLRDXYTRYVL 335  
DB VEENSMYDPTIYEDFLTHIFIONQOEVYTKLLHMFDELOVYAHKTOELRDVTRYVL 355  
QY 336 WYNNHFNDEGDPAMTRFLEEFENNLEREKMGHLRLNIAACAARLRLMTLTKPSREA 395  
DB WYNNHFTDFADYEMMEFEVFEALERKLLQLRLHIAACAARMSCTLTSSRDE 415  
QY 396 ELPEILLGSE-----KGFIFVDVSDGSKATEGLRGOILEVNOQNEINTQ 445  
DB PLNFRTYGVGELRGVALTGNAAVGIYISHVEGSAQDGLRGQIHEVNOQSDIHYT 475  
QY 446 LSKAMEILRNTHLSITVKNLFFVKELLTRLSEE-----KRNKAGPLP 489  
DB SKRALELTCTHLSISVKNLGLFEKIMQALHGGGTAGSGSISAGSGFKVRSRRT 535  
QY 490 KIDIRK-----AS-----RYSIPDLAVD----- 508  
DB CANDIAKLHGRSDTDELSSVASNRAHMYRLSSVMDLDDPCAPQTPPVSGSNMA 595  
QY 509 ---VEQYIGLEKN---KSKAN-----YGG-----RKLKILDKTISTLPK 548  
DB SNFMQOL--LQSVNNSAKKSGNSNSDDQDTKGGEFTLAPKRRLQKALAKML----- 647  
QY 549 PYNDIGIGGODSDIVGL-----RQTKHPIPTALPVSG---TLSSNPDL 589  
DB LKQNHGSSLNDSDTLNDPCKSLSAVSSCSSTOSSINGCTVSGGGLYIOSOSNPDL 706  
QY 590 -----LQSH-HR-----ILDFSAT-----PDLPOVLERYKA 615  
DB TSLNYDGGSDAGNGGGRLOVYLNIAHHRPASAFTLTNSTOSHLLPDYDPDVLKYKA 766  
QY 616 DOOSRYIMISKDTAKENVYQAIRERAVYATPPOQYSLCEVSVPEVYIKRRLPDOLSKL 675  
DB DOCKYVLYLKETTAHEVALTLQIEFGIHPSSNFSICEVSDGCVKORRRLPDOLNL 826  
QY 676 ADRILOISGRVYLLKNMTEFLCDEDAQOELLRESOISLQLSIVEVATOLSMNFELEFRN 735  
DB AERISPAARKYLLANDSTELVDELALBELVRSNVHFLNLYELAIOTLDDPGFGRQ 866  
QY 736 IEPTEYIDDLFKLSKTSKANLRFEEVINOETFWVASEILRETNOLKRAKIIKHFIXIA 795  
DB IESTEVDELFEELRSRGVMLSKPAELVREMFWVYSELCAHNYVRMKIKYKOFIXIA 946  
QY 796 LHRBECKNFMSALISGLNALAVARLRTWEKLPNKYEKLFODLDDLPSSNMAYRN 855  
DB RHCKEGRNFMSPAIVSIGLGHGAVSRRLQRTWEKLPKRYKQLFNDLDDLPSSNMAYRN 1006

QY 856 VLNSQNL-OPPIPLRPVIAKIDTLFLEHGNDSKVDGLVNFEEKLRMTAKETIRHVGMAVYN 914  
DB LVSABELLAQHPIIPFPYIKKDLTFHLGNDTPVDLDLINEKRLMAKAVRLLTHMCCSP 1066  
QY 915 MD---PALMFTRRKKRWSLGLSOGSTNA---TVLDVAQTGH---KKVARRSSFLLNA 964  
DB YDLSTLELTKGQSPNALFSLNMSASQSNAAAGTV--IAANAGQATIKRRKKSTAAPNP 1124  
QY 965 KRIYEDAOAKARKYKOLYLSNLELEMEESLQTLSDCEPA-----TNLPK----- 1009  
DB KMFEEAOVVRKAVYLNLSIKLILSDLELHFKLSLECEPAGSGTYSGSISHVNTSHRSGG 1184  
QY 1010 -----NPDKKPVKSETPSPAPRAGSOOKASOLPQOQOPPAH 1048  
DB GSISGAGSGSGGGGGSSSLNAGDQSLTSHTS-----SSAPNSSLKRRHP----- 1234  
QY 1049 KINQGLQVPAVSLYPS-----RKVPYKDLPPFGINSQALKILSLSESGLERHKK 1101  
DB -----SSPFLSTSTSTSTSDHQROMHNNCPKFGTASPOAVKMLSLSESKIRPH-- 1286  
QY 1102 QAEDITSMASQSL--PPSPSSPRKGYTLAPSGTVNDFSDSGHSEISRSSTYSNS 1159  
DB -----QPEVPRHGSTMAVYIPPLHMHAAHGFSTPSPGVV-----TSPATSAVAAWQ 1334  
QY 1160 FDSVPVSLHDERQRHSVSIVETNLGMRERRTMIEPD-----QXSLGYPAMSEGRGL 1214  
DB CTSPSP-SPCHRLRLAAGNITIPRAIHERSHSTPAPRPLPVSVDLSLE----- 1383  
QY 1215 YATATVYISFSTEELSODQDGRASLDAADSGRSWTSKSSGSHDNIQTIOHRSWETLP 1274  
DB -----SSVTT---FRDPLRKSVTs---GSISCDG-----YVHQOQ----- 1415  
QY 1275 GHTHPDYSGDPALMASSSIMDQIMFSDHSTKYNRONQSESLQOASRMASSSTGYWG 1334  
DB 1416 -HYHLQY-----QOQOQOQNSOHEPSPPVYTA----- 1442  
QY 1335 EDESGDTGTRKRGCKDVSILEASSLTSVTEETKPVMP-----AHIAVASTTGCLA 1390  
DB 1443 -----DORLLQOISNNAVYTNL-NSPCQSTNTPPTPTPPNPOTATIQASAPTAAYM 1496  
QY 1391 RKEGRYRE-----PPPTPGYIGIPITDPPEG--HSHPARKPPDYNAVLAORSRMV 1438  
DB 1497 HARSQHOQLOQOQOOSLAMPPEPPRY-NVP-----PLGSIVSH----- 1533  
QY 1439 ARSSDTAGPSSVOQPHGPT-----SSRPVANKPQ 1467  
DB 1534 --HQGTAGSRHLNMHGKTTGPOERWFPDRCPTTKQO 1568  
RESULT 10  
O43386  
ID O43386 PRELIMINARY; PRT; 579 AA.  
AC O43386;  
DT 01-JUN-1998 (Tremblrel. 06, Created)  
DT 01-JUN-1998 (Tremblrel. 06, last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
DE KIA001LB (Fragment).  
GN KIA001LB.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Connolly K.S., Gunning K.M., Davis C.A., Kadner K., Subramanian S.,  
RA Miguel T., Lewis K.D., Fridlyand J., Alciatore D., Benke J.A.,  
RA Bondoc M., Bowen E., Chiang A., Critz P., Jaklivic M.A., Lindo K.,  
RA Lindquist K., Miller C., Patel S., Piscià C., Riley B.E., Rojeski H.,  
RA Sarmiento R., Yu C., Montenegro M., Aerts A., Chung A., Adriano A.,  
RA Baker M., Gau C., Jett J., Ko C., Beall K., Woolley J.P., Stultz J.L.,  
RA Kimerly W., Martin C.H.;  
RT \*Sequencing of human chromosome 5.\*;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

RN [2]  
SEQUENCE FROM N.A.  
RP Riecke D.O., Wagner R.P.:  
RT "Large Scale Sequence Analysis and Annotation with the Sequence  
Comparison Analysis (SCAN) System."  
RL Submitted (Feb-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC004227; AAC04379.1; -  
DR HSSP: Q12923; 3PDZ.  
DR InterPro: IPR000595; CNMP\_binding.  
DR InterPro: IPR001478; PDZ.  
DR InterPro: IPR000651; RasGEFN.  
DR Pfam: PF00027; CNMP\_binding; 1.  
DR Pfam: PF00595; PDZ; 1.  
DR Pfam: PF00618; RasGEFN; 1.  
DR SMART: SM00100; CNMP; 1.  
DR SMART: SM00228; PDZ; 1.  
DR SMART: SM00229; RasGEFN; 1.  
DR PROSITE: PS50042; CNMP\_BINDING\_3; 1.  
DR PROSITE: PS50106; PDZ; 1.  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 579 AA; 65197 MW; CAE221DE69CBF92D CRC64;

Query Match 28.1%; Score 2173; DB 4; Length 579;  
Best Local Similarity 73.0%; Pred. No. 2e-117;  
Matches 425; Conservative 70; Mismatches 79; Indels 8; Gaps 7;

QY 21 LPADFTKLTHTDSLHPQVTVSSSHSGCSITSDSGSSLSDIYQATSEAGNDMLSLPE 80  
DB 1 LPADLTLMHLEPHVTVSSSSQCSIASDSSSLSDIYQATSEAGNDMLSLPE 60  
QY 81 TAVDSEDDDE-EDIERASDPLMSRDIVRCLEKDPIDRTDDDEQLLEFMHQLPAPAM 139  
DB 61 GPVDSDEDEDEDEIDR-TPLQGRDLVRCLEKPERADKDDDEQLLEFMHQLPAPAM 119  
QY 140 TMSYRRELCAVMFAVERAGTIVLNDGEELDSVILNGSVLEVYTPDGKAEILCMGNSF 199  
DB 120 TMSYRRELCAVMFEVEYQAGAILLEDGQELDSVYILNGTVEISHPDGKVENLFGNSF 179  
QY 200 GVSPTMKETKMGVMTKVDCCQVCAQDQYCRILNQVERKMGQVEEKEIVVVKHRE 259  
DB 180 GTPPTLKQIMHGIVKRVKVDCCQVCAQDQYCRILNQVERKMGQVEEKEIVVVKHRE 239  
QY 260 LDRGTGRKHIVIKGTSERLTMLVEHSHVVDPTFIEDFLTYRTPLSSPMYEGKLLW 319  
DB 240 LDRSGTRKHIVIKATPERLIMLIEHSHVVDPTFIEDFLTYRTPLSSPMYEGKLLW 299  
QY 320 FNDPSLDKTVTVLVVNNHNFDEBGPAMTRLEFEENLREKMGHRLNLTA 379  
DB 300 FKIDSLDKTVTVLVVNNHNFDEBGPAMTRLEFEENLREKMGHRLNLTA 359  
QY 380 KAKRLMTLTPREAPLPFILLGSEKSGCIFVDSVDSKATEAGLKGDDILEVNGQ 439  
DB 360 KAKRQVYVLOKASRESPLQSLNGSGKSGCIFVEGEPESKADSLSKAGDDIMEVNGQ 419  
QY 440 NFENIOLSKAMEILRNTHLSTIVKTNLFVFKELLRLSEKRNAGRLIGDIKASR 499  
DB 420 NFENITFMKAVELIRNTHLSTIVKTNM--FKELLFR-TEQEKSGVHPHIAE-KKSNR 475  
QY 500 YSLPDLAVDVOYIGLEKVKKSKANTV-GGRKKLKKILDKTRISILPQRPYNDIGIGOS 558  
DB 476 HSIQHPGDEQOT-SQEKSGKVKANPVSQGRNKRIRKILDKTRISILPRLFSQGLSOS 534  
QY 559 ODDSIIVGRLTKHLPALPVSGTSSSNPDLQSHHRIIDFS 600  
DB 535 QDDSIIVGTRHCRHSLAIPGLTSSSPDLQPTTSMIDFS 576

RESULT 11  
Q95WR8 PRELIMINARY; PRT; 1347 AA.  
AC Q95WR8;  
DT 01-DEC-2001 (Tremblrel. 19, Created)

DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Pxf isoform C.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabditoidea;  
OC Rhabdilitae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxId=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Verheljen M., van Berkel W., Jansen G., de Rooij J., Plasterk R.H.,  
RA Bos J.L., Zwartkruis F.J.T.;  
RT "Characterization of pxf, the C. elegans homolog of human PDZ-GFPs."  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF308449; AAU09435.1; -  
DR InterPro: IPR000595; CNMP\_binding.  
DR InterPro: IPR001478; PDZ.  
DR InterPro: IPR000651; RasGEFN.  
DR InterPro: IPR001895; RasGRF\_CDC25.  
DR InterPro: IPR00159; RA\_domain.  
DR Pfam: PF00027; CNMP\_binding; 1.  
DR Pfam: PF00595; PDZ; 1.  
DR Pfam: PF00617; RasGEF; 1.  
DR PROSITE: PS50042; CNMP\_BINDING\_3; 1.  
DR PROSITE: PS50106; PDZ; 1.  
SQ SEQUENCE 1347 AA; 149268 MW; 19708909E535BC0E CRC64;

Query Match 23.3%; Score 1798.5; DB 5; Length 1347;  
Best Local Similarity 35.3%; Pred. No. 3.5e-95;  
Matches 459; Conservative 216; Mismatches 435; Indels 191; Gaps 36;

QY 80 ETVVDSDEDDDEED-----TERASDPLMS-RDIVRCLEKDPIDRTDDDEQLLEFMHQL 133  
DB 163 ETVSNAGADEDEDEGSMPSQESSGGFMRLDSVRELEKEPERKSESDLAVIDPMQHW 222  
QY 134 PARANMTSVYRRELCAVMFAVERAGTIVLNDGEELDSVILNGSVLEVYTPDGKAEIL 193  
DB 223 SAFAALPMSIKRQLCKMFAVYVNDAGTVLANNEKLDMSVYVNCQVEVVKSGRVEV 282  
QY 194 CMGNSFVSPTMKETKMGVMTKVDCCQVCAQDQYCRILNQVERKMGQVEE-EGEL 251  
DB 283 KIDDSFAEPTPATQIHIGEMRTMVDCEFLVLEHNFCSIMSTIDHIEKDRDGLTGEV 342  
QY 252 VMKHEHRELDRTGRKHIVIKGTSERLTMLVEE--HSVVDPTFIEDFLTYRTPLSSP 309  
DB 343 VSEVERKTV---GTHCGQVILKRPKRLNHLVDEKDN-VDRHYVDDFLTYRTVIRRP 398  
QY 310 MEVGGKLLMFNDPSLDKTVTVLVVNNHNFDEBGPAMTRLEFEENLREKMGH 369  
DB 399 TTFTEKIMLFAFSIYRDKVAVRLVLLVNNHNFDEFTNDEMNNLLERFGALERDGMHSQ 458  
QY 370 LRLNLTAACAARRLMTLTPREAPLPFILLGSEKSGCIFVDSVDSKATEAGLKR 429  
DB 459 LSLNLTAACSVKAPROYVILR-RKDDKMMRLVGGESGVSVAVFPDTSARREGVR 517  
QY 430 GDOILEVNGQNFENIOLSKAMEILRNTHLSTIVKTNLFVFKELLRLSEK-----FN 483  
DB 518 ADMELEVNQSAKYLLAKKAEDELTSLSLTMLKNNVLGKRTICKIEHNRKKNQTSNS 577  
QY 484 GAPHLRIGDIKASRYSIPDLAVDVOYIGLEKVKKSKANTVGGRNKLKLT----- 537  
DB 578 GA-GIPMVPIPVHKT-----ITGKKSSTTSKS---GMMEKLMTLKSKED 620  
QY 538 -----DKTRISILPQRPYNDIGIGOSODDSIVGLRQTKHLPALPVSGTSSSNPDLQS 592  
DB 621 SMDFTDEAKISSADLRP-----SRSNPDITS- 646  
QY 593 HHRILDF--SAPDLPDVOYLVEFKADQSRVIMISKDTAKEVYIOAIRFPAVTAT-PDQ 649  
DB 647 ---ISQYGPVRSCEPHYKLTIRNDQTKRYLVYKETSQNVQALDELFFNTTAGSGPE 703  
QY 650 YSLCEVSVTPEGVTKQRRLPDQLSKLADRIQLSGRYLLKNNMETETLCDEDAQOELLRES 709

```

Db 704 WSICECTVITIDVIVIKORRLPPOMENIALERIALNSRYILKNNRSRSEPLVPDELAPELLKEA 763
QY 710 QSLSLQSLVEVATOLSMNPELFNRIEPTXIDDLFLKRSCTSCANKKREFEVNOEFF 769
Db 764 QQLSLNQAQVAAQQLTLQDFSVFSAIEPTERLQDLFKDSYSGPKLEEFQQLNREKW 823
QY 770 WVAISELRETNOLKRRKIKHFIKHTALHCRECKNPNSEFAISGLNLAVARLRTWEXL 829
Db 824 WVAATEICFERHQAQKAKLIKFKFIVARCRDLRNFNSEFAMISGLDKPAVRRLHSWEV 883
QY 830 PAKYKLELQDLODLDPFSNNAKYKRVNLSNOLQPIIPLFVYIKKDLTLEHGDSKVD 889
Db 884 SSKYIRMEDEIHQILVDPSPKNSMKRQHLAEVQEPVPVPIYVIKKDLTFADGNATYSE 943
QY 890 GLVNFETKLMIAKEIRHYGRMASVNMDDPALMFRTRKKKRSISGSGSTNATVLDVQCT 949
Db 944 KLINEKRLAKRSKRWKGLSSAPYETASMAE-----RSGGVMDALLHNNSFENSRY 997
QY 950 GGHKRRVRSSEFLNAKKLYEDAQMARKVQYLSNLELEMEDESLQTLSTOCEPATNTLPK 1009
Db 998 ATMKGMSGKQNPQPKKYEQALMVRKYKSYLEGLHYVDNEMELDSMSYDIEPOQTARH 1057
QY 1010 --NPDCKKPVKSETSPVAPRAGSQKQKOSLPQPOQPPAKH---NOGLQVPAVSYPY 1064
Db 1058 GANSSSTANIRRVSPSP--TPSSLSOSAGSADQSSR-----HRLTFNGTG---SISAG 1107
QY 1065 RKKVYVVKDLPFGINSPOAKKIILSISEGSLERHKKQAEITISNASSLSPPSPSS 1124
Db 1108 GSK-----FVESPQAVQKMLSLVQNSK-----VKAPPOITSPSTARS 1148
QY 1125 -----PRKGYTLAPS-----GTVDNF--SDSGSEISSRSSIVSNSSPDSVY 1165
Db 1149 LQNNPRTVGRATTSAGCPVQVLNETSYTYTYOSDNKRORSQ-----SGCRDNIP 1203
QY 1166 S---LHDER---RQHSVSVETNLGCMGRERTMIEPQYSLGSAVPSSEGRGLYATAT 1219
Db 1204 STFYLTSDGLTVSPROSLSVVIPTHPHGH-----SPTSPRCRSRSPASSCSFSSTA 1256
QY 1220 VSSPSTEELSDQDQDRASLDADSGRGSWTSCSSGSHNIOITQ-HQSMWETLPF---G 1275
Db 1257 STAA-----TSMAPAAPS-----AFVSNPYOHOHTVGHVIGHRPMPIVTS 1297
QY 1276 HTHFDYSGDPAGLIMASSHMDQIMFSDHSTKYRONQSR 1316
Db 1298 SATLPNHNVSRLPPKS--RPTILPGSHNSSRMGTIKEA 1336

RESULT 12
Q21218 PRELIMINARY: PRT: 1470 AA.
ID 021218: Q22503: Q9U4J1:
AC 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE T14G10.2A protein (RA-GEF) (PXF Isoform A).
GN T14G10.2A OR RA-GEF.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN NCBI_TaxID=6239;
RA none;
RP SEQUENCE FROM N.A.
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RL investigating biology."
RL Science 282:2012-2018(1998).
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-BRISTOL NZ:
RX MEDLINE=20076450; PubMed=10608844;
RA Liao Y., Kariya K., Hu C.-D., Shibatahge M., Goshima M., Okada T.,
RA Watarai Y., Gao X., Jin T.-G., Yamawaki Katsoka Y., Katsoka T.,
RT "RAGEF, a Novel Rap1a Guanine Nucleotide Exchange Factor Containing a
RT Ras/Rap1a-associating Domain, Is Conserved Between Nematode and
RT Humans."
RN J. Biol. Chem. 274:37815-37820(1999).
[4]
RP SEQUENCE FROM N.A.
RA Verheljen M., van Berkel W., Jansen G., de Rooij J., Plasterk R.H.,
RA Bos J.L., Zwartkruis F.J.T.,
RT "Characterization of pxf, the C. elegans homolog of human PDZ-GEFs."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z68880; CA93100.2; -.
DR EMBL: Z69664; CA93100.2; JOINED.
DR EMBL: Z69664; CA93519.2; -.
DR EMBL: AF107956; AAF22963.1; -.
DR EMBL: AF308447; AAL09433.1; -.
DR InterPro: IPR000595; CNMP_binding.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR000651; RasGEFN.
DR InterPro: IPR001895; RasGEF_CDC25.
DR InterPro: IPR000159; RA_domain.
DR Pfam: PF000027; CNMP_binding; 2.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00788; RA; 1.
DR Pfam: PF00617; RasGEF; 1.
DR SMART: SM00100; CNMP; 1.
DR SMART: SM00228; PDZ; 1.
DR SMART: SM00314; RA; 1.
DR SMART: SM00147; RasGEF; 1.
DR SMART: SM00229; RasGEFN; 1.
DR PROSITE: PS50042; CNMP_BINDING_3; 1.
DR PROSITE: PS50106; PDZ; 1.
SQ SEQUENCE 1470 AA; 163807 MW; 59404704C1346BC8 CRC64;

Query Match 23.3%; Score 1798.5; DB 5; Length 1470;
Best Local Similarity 35.3%; Pred. No. 4e-95;
Matches 459; Conservative 216; Mismatches 435; Indels 191; Gaps 36;

QY 80 ETAVSEDDDEDEDE-----IERASDPLMS-RDVIKRCLEKPDIDRTDDIEQLLEPMHQL 133
Db 286 ETASNEGADSDDEDEGSMPSQESSGCFMDLRDSVRCLEKESERNSEDLAVLDMQNM 345
QY 134 PAFAMTMSVRRLECAVYFAVERAGTIVLNDGEBLDSWYILNGSVETYPDGKAEIL 193
Db 346 SAFALPMSIKRQQLCKMFAVAVNDAGTVLVAHNEKLDWSYIVNGCVAVRPSGRIEY 405
QY 194 CMGNSFGVSPYMDKEYMKCVMTKYDDCOFVCIAOODYCRILNOYEKKNQKVEE--EGEI 251
Db 406 KLGDSGAGPPTPATQIHIGEMRTWDDCEFLVEHRDFCSIMSTIGDHIEKDRDGLTGEV 465
QY 252 VAWKEHRELDRGTGRKHIVIKGTSERLMLHVEE--HSVVDPTFIEDLLTYRFTLSP 309
Db 466 VSEVERRY--GTHCGVLLKIGKPKLHLHLYDERDHN-VDPHYVDLTLTYRFTIRDP 521
QY 310 MEVGRKLLMEFNDPSLRDQVTRVVLVWNNHFNDEGDPAKRFLEEFNNLEERKMGH 369
Db 522 TTFIEKLMFMFADSIYRDQVAVRLVLLVWNNHFNDEFTNDEMNNLEERFGALERDMHQ 581
QY 370 LRLNLACAKAKKRLMTLTKPSREAPLFPILLGSEKFGILFVDSVDSGKATEGLKR 429
Db 582 LSLNLACSVKAKKPROVILTR--RKDDKMMKRLVCGQESGNSYVAVEVPDTSAAREGYK 640
QY 430 GQOILVNGQNFENIQLSAAEILNNTLSTIVTNLTFVFEKLLRLSEEK-----RN 483
Db 641 ADEMLEVNOOSAKYLSAKKAEIDLTGSLSTLMLKNVLYGETIGKLEHNPKNGTSS 700
QY 484 GAPHLPKIGDIIKASRYSIDPLAVVEQVIGLEKYKKSKATVGGRNKLKIL----- 537
Db 701 GA-GIPMTVIPVAKTS-----ITGKKSSSTSSKS---GMMEKLMITILSSKED 743

```





[illegible]

```
Db      1 TNSNMADV-OGGAHKRRARRSSLLMAKKLYEDAOMARKVKYLISSLIDTIDEKFOHMSL 59
Qy      999 OCEPPTNLTPENPPDKRPVK-SETSPAPRAGSQOKAOSLPPOQQOPPAKHINGLOVP 1057
        ||| ||| ||| :|| :|| ||| ||| | ||| ||| :||| |||
Db      60 QMEPYGLTTLNLTERRSAKSSEKSPPLRLSVQOTAKVHLHQD-----HRYSOVLDP 112
Qy      1058 AVSLYPSRKVPVVDLDPFGINSFQALKILSIEGSELRHKKQAEDYTISNASOLSPP 1117
        ||| ||| ||| :|| :|| ||| ||| | ||| ||| :||| |||
Db      113 AVNLHPRIKKGOARD----HYLSLTPQKGCLGPTEEVSXKH---TEDITIVASSLSHSP 165
Qy      1118 PTFQSSSRKGYTLAPBGTYDNSTDSGHSTLSRSSITSVNSSFDVYPVSLHDERORHSV 1177
        ||| ||| ||| :|| :|| ||| ||| | ||| ||| :||| |||
Db      166 PASPNQSNPKRYTTLTPSSKCNDLSDSSHSEISSRSITSNGSVDSMAGAODERCSSHL 225
Qy      1178 SIVENIIMGMERMTIMEPDQYSLSGSYAMSGRGITYATAYTSPTFEELSQQGDRA 1237
        ::||::||::||::||::||::||::||::||::||::||::||:
Db      226 AVEPTGALETHDPBSGI--SDHSQLAHGMWLSKPCLINGVAVSSLSSEEMSH---EHV 280
Qy      1238 SLDAADSRGSWTSCSSGSHDNIOTIOHORSWETL-PFGTHFDYSGDPAGLMASSSHMD 1299
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      281 VLEAADSRGSWTSCSSSHDHQSDLNOKSMDFLSTRMHLD---DT---AEPEYPD 334
Qy      1297 QIMFSHSTKYNRONQRESILEQASPRASMASSTGYWGDESEDGTIKIRKGKDVSIEA 1355
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      335 CEPCCAPKRCGRTOCGCKLETUOLQRWASSSSL-SDTCBNPTQVARKYLESPAWEA 393
Qy      1337 ESSSLTSTYTEETFPVMFAHIAVASTTGCLA-----RKRGRTREPPPIPGYIGIP 1410
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      394 PDGLEPRTUD-----PVKYTVSTSDGLLVYCVTSPKKDRKRREPPPIPGYLGIS 446
Qy      1411 ITDFEGHSHARPDPDYNALQGRMYAVARSSDTAGPSSVOQPHGPTS---SRPVNKPO 1467
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      447 LADJKEG-PHHKLKPPRYSAVAVQSKMKMLSLERLPPA---PPSSHTSAMVSKSIGQPQ 502
Qy      1468 WKPNESDPLARYQSQGFTSEDEDEQVSAY 1499
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      503 RH---SHPKLADYADA-DSEADENEQVSAY 528

RESULT 15
OBVCC8 PRELIMINARY; PRT; 876 AA.
AC OBVCC8;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Similar to cAMP-regulated guanine nucleotide exchange factor I
   (cAMP-GEFI).
DE OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/Genbank/DDBJ databases.
DR EMBL; BC020532; AH420532.1; -.
DR InterPro: IPRO002373; CAMP_kin.
DR InterPro: IPRO000595; CNMP_binding.
DR InterPro: IPRO00591; DEP.
DR InterPro: IPRO00651; RasGEFN.
DR InterPro: IPRO01895; RasGRF_CDC25.
DR Pfam; PF00027; CNMP_binding; 1.
DR Pfam; PF00610; DEP; 1.
DR Pfam; PF00617; RasGEF; 1.
DR Pfam; PF00618; RasGEFN; 1.
DR PRINTS; PR00103; CAMPKINSE.
DR SMART; SM00100; CNMP; 1.
DR SMART; SM00049; DEP; 1.
DR SMART; SM00147; RasGEF; 1.
DR SMART; SM00229; RasGEFN; 1.
DR PROSITE; PSS0042; CNMP_BINDING_3_1.
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2003, 17:42:44 ; Search time 41 Seconds

(Without alignments)  
1516.415 Million cell updates/sec

Title: US-09-911-826A-2

Perfect score: 7721

Sequence: 1 MKPLAIPANHVGMQGEKHS.....PYQSGFSTEDDEDEQVSAY 1499

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SWISSProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	533.5	6.9	1234	1 YNXX_CAEEL	P34578 caenorhabd
2	343	4.4	1595	1 SOS_DROME	P26675 saccharomy
3	325.5	4.2	1095	1 CC25_SACKL	002342 saccharomyc
4	314.5	4.1	1253	1 SC25_YEAST	P14771 saccharomyc
5	309	4.0	1262	1 GNRP_MOUSE	P27671 mus musculu
6	307.5	4.0	1319	1 SOS1_MOUSE	062245 mus musculu
7	303.5	3.9	1333	1 SOS1_HUMAN	Q07889 homo sapien
8	303	3.9	1333	1 CC25_CANAL	P43069 candida alb
9	290	3.8	1275	1 GNRP_HUMAN	013872 homo sapien
10	288	3.7	1244	1 GNRP_RAT	P28818 rattus norv
11	282.5	3.7	1589	1 CC25_YEAST	P04821 saccharomyc
12	277	3.6	911	1 STE6_SCHPO	P26674 schizosacch
13	245.5	3.2	1297	1 SOS2_MOUSE	Q02384 mus musculu
14	244.5	3.2	768	1 RGL1_HUMAN	Q9a216 homo sapien
15	242.5	3.1	1332	1 SOS2_HUMAN	Q07890 homo sapien
16	238.5	3.1	768	1 RGL1_MOUSE	060695 mus musculu
17	227	3.1	2349	1 TPR_HUMAN	P12270 homo sapien
18	236.5	3.1	852	1 GND5_MOUSE	Q03385 mus musculu
19	234.5	3.0	914	1 GND5_HUMAN	Q12967 homo sapien
20	232	3.0	1435	1 LTEL1_YEAST	P07866 saccharomyc
21	229.5	3.0	895	1 GND5_RAT	Q03386 rattus norv
22	229.5	3.0	1745	1 ZOI1_MOUSE	P39447 mus musculu
23	218	2.8	1736	1 ZOI1_HUMAN	Q07157 homo sapien
24	217.5	2.8	777	1 RGL2_HUMAN	Q15211 homo sapien
25	211.5	2.7	778	1 RGL2_MOUSE	Q61195 saccharomyc
26	211	2.7	1875	1 MLP1_YEAST	002455 saccharomyc
27	206.5	2.7	2845	1 APC_MOUSE	061315 mus musculu
28	206.5	2.7	3122	1 DPOZ_MOUSE	061493 mus musculu
29	200	2.6	1111	1 KIP1_YEAST	P28742 saccharomyc
30	198.5	2.6	3210	1 KLEF_HUMAN	P49654 homo sapien
31	197	2.6	1066	1 KLE1_DROME	P46863 drosophila
32	196	2.5	2464	1 MABP_MOUSE	P14873 mus musculu
33	193.5	2.5	1953	1 BNT1_YEAST	P41832 saccharomyc

34	191.5	2.5	1184	1 B1MC_EMENT	P17120 emericella
35	190.5	2.5	2459	1 MABP_RAT	P15205 rattus norv
36	190	2.5	2231	1 SEN1_YEAST	Q00416 saccharomyc
37	189	2.4	1462	1 NKCR_HUMAN	P30414 homo sapien
38	189	2.4	1727	1 ALM1_SCHPO	Q9utk5 schizosacch
39	187.5	2.4	2145	1 CYAA_PODAN	Q01513 podospora a
40	187	2.4	2472	1 NCR2_MOUSE	Q9w42 mus musculu
41	186.5	2.4	897	1 EP15_MOUSE	P42567 mus musculu
42	185.5	2.4	3924	1 ANK2_HUMAN	Q01484 homo sapien
43	184.5	2.4	538	1 BUD5_YEAST	P25300 saccharomyc
44	184.5	2.4	1957	1 YD86_SCHPO	Q10411 schizosacch
45	184.5	2.4	2842	1 APC_RAT	P70478 rattus norv

## ALIGNMENTS

```

RESULT 1
ID      YNXX_CAEEL      STANDARD:      PRT: 1234 AA.
AC      P34578;
DT      01-FEB-1994 (Rel. 28, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Hypothetical protein T20G5.5 in chromosome III.
GN      T20G5.5
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Rhabditidae; Peloderinae; Caenorhabditis.
OX      NCBI_Taxid=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Bristol N2;
RA      Berks M., Smith A.;
RL      Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.
RN      [2]
RP      REVISIONS.
RA      Durbin R.;
RL      Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
CC      -1- SIMILARITY: CONTAINS 1 DEP DOMAIN.
CC      -1- SIMILARITY: CONTAINS 2 CYCLIC NUCLEOTIDE-BINDING DOMAINS.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
-----
CC      EMBL; Z30423; CAAB3013.2; -
CC      PIR; S42368; S42368.
CC      HSSP; P00515; 2BPk.
DR      Wormpep; T20G5.5; CE23992.
DR      InterPro; IPR000591; DEP.
DR      InterPro; IPR000651; RasGEFN.
DR      InterPro; IPR001895; RasGRF_CDC25.
DR      InterPro; IPR002373; CAMP_kin.
DR      InterPro; IPR000595; CAMP_binding.
DR      Pfam; PF00027; CNMP_binding.
DR      Pfam; PF00610; DEP_1.
DR      Pfam; PF00617; RasGEF_1.
DR      Pfam; PF00618; RasGEFN_1.
DR      PRINTS; PR00103; CAMPKINASE.
DR      SMART; SM00049; DEP_1.
DR      SMART; SM00229; RasGEFN_1.
DR      SMART; SM00100; CNMP_2.
DR      PROSITE; PS00888; CNMP_BINDING_1; FALSE_NEG.
DR      PROSITE; PS00889; CNMP_BINDING_2; FALSE_NEG.
DR      PROSITE; PS50042; CNMP_BINDING_3; 2.
DR      PROSITE; PS50186; DEP_1.
KW      Hypothetical protein; Repeat.
FT      DOMAIN 206 336 CNMP-BINDING 1.

```



FT	DOMAIN	247	433	DH.
FT	DOMAIN	479	587	PH.
FT	DOMAIN	1511	1516	GLN-RICH.
FT	DOMAIN	1225	1541	HIS-RICH.
FT	CONFLICT	232	243	TSCVPCCHFPSS -> HILPSPSLPAOR (IN REF. 2).
FT	CONFLICT	1462	1462	V -> P (IN REF. 2).
SO	SEQUENCE	1595 AA;	177837 MW;	33AE31F0767A219F CRC64;
Query Match				
Best Local Similarity		4.4%;	Score 343;	DB 1: Length 1595;
Matches 298; Conservative 195;		17.1%;	Pred. 1.6e-08;	
			Mismatches 452;	Indels 794; Gaps 61;

[illegible]

QY	760	FEEVYNQETTFWAVASEILREINQKRMKIIHFKILALHCECKENFMSFALLISGLNLAIV	815
Db	872	IMKFTTIVATRWIKESITDEAEYERLAIMORALEVMKVMLENNFGLISTIVAAMGTASV	931
QY	820	ARLRTEWTEKLPNKYEKTLFODJQDFDSRYMAKYRVNVLNSQNLQPPILPFPYIKDITLF	879
Db	932	YRLAMTFQGLPERKRLLEBDRSDQ--HLKKYQELRSIN--PCVPRFFGKYLNIILH	987
QY	880	LHESGDSKYDG--LVNFEKLMIAKETIRHVGMAVSVMDBALMFRTKKRKWRSLSLSOG	937
Db	988	LEENPDLANTLLETINFSKRKVAEIT-----GEIQY	1022
QY	938	STNATVLDVADQGTGKKRVRSSLTNAKKLYE-----DAQMAKKQYILSNLELM	988
Db	1021	QNPQYCNLEEST-----TROFEEQLDPFNGLSIDQMS---DYLYNNESLRI	1067
QY	989	DEESLQTLSTLQCEPATATLPLKPNPGDK-----KPVKSEFSPVAPRAGSQOKQASLP	1033
Db	1063	EPRCCKTIVPRPKRWPIPLKSPQIKRRRONQTNSSKKLSNSTSSVAAAAMASSTATST-	1122
QY	1039	QPPQOPPRAHKINQGLQVPAVSVLYPSRKKRVKVDLPFGINSQALKKILSLEESLER	1099
Db	1122	-----ATASAPLSLHASINDAP-----	1133
QY	1099	HKQQAEDTINASSQLSPPTSPQSSPRKQITLAP-----SGTYDNF--SDSGHS	1144
Db	1139	-----TAAANAGSGTLAGEOSDPHNHAFSVAPVPIPERNTSSMSGPQHFRTIQDNG	119
QY	1147	EIS-----SRSSIVNSSEFDSV-----PVSIDERRQRRISVI	117
Db	1194	EVSYPAPHLPRKPGAHWANNNSTLASASMDVFFSPALDEHLRPOSLPDPSNFPASDTGA	125
QY	1180	VEINLGMGRMERTMIEPDQYSLGSYAPMEEGKGLYATATVIVSSPSTELSDQDGRASL	1233
Db	1254	PPSP-L-----PKLVASPR--RHETGNRSPF--HGRMQNSPTHTASTAVTTL	129
QY	1240	DADDSGKGSWTSCSGSHDNITQIHQRSMETLPG-----HTHEBDYSGDPAGL----	1287
Db	1295	--GMSTSGGEFFCGGGEYFN---SAHQGGPGAVPISPHVNVPMATNMEXKAVPRPLPRR	134
QY	1289	---WASSSHMQOIMES-----DHST--KYNRQMSHESLEQAO	132
Db	1350	KERTESCDAQAQKQADAPDAPLPPRQSGELSPPTLPPALNSTGISTYLROSHGK-----	140
QY	1322	SRASWASSTGYWGEDS--EGDGTGIRKRGKDVSIIEAESS-----LT	136
Db	1403	-----SKEVGNSSLLPLPTSSIMIRRSALIEKRAALISQGNQAAGPISITLVTVSQ	145
QY	1333	SVTTEETKPVDM--PAITAIVASSTTKGLIAREKGRYREPRPTPGYIGIPITDFECSHP	1422
Db	1456	AVATIDEVLPILPISPA--ASSSTTSPPLTPAMS-----PMSNPINPSHPVESTSSYAH-	150
QY	1422	ARKRPDYVVALQBSRMVARSSDTAPGSVVOQRGHRTSSRPVKKRPMQHNKPNEDPLAR	1480
Db	1506	-----QLMRQOQOQOQTNPATLYISQHNQHNATHLPR--HHPRQHNHSNPQSRSP	1551
RESULT 3			
CC25_SACKL			
AC	002342	STANDARD;	PRT; 1095 AA.
DT	01-JUL-1993	(Rel. 26, Created)	
DT	01-JUL-1993	(Rel. 26, Last sequence update)	
DT	15-JUN-2002	(Rel. 41, Last annotation update)	
DE	Cell division control protein 25 (Fragment).		
GN	CDC25.		
OS	Saccharomyces kluyveri (Yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
OX	NCBI_TaxID=4934.		
FN	[1]		
RP	SEQUENCE FROM N.A.		
FX	MEDLINE=92354938; PubMed=1644315;		

```

RESULT 3
CC25_SACKL
ID      CC25_SACKL      STANDARD;      PRT;      1095 AA.
AC      Q02342;
DT      01-JUL-1993 (Rel. 26, Created)
DT      01-JUL-1993 (Rel. 26, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Cell division control protein 25 (Fragment).
GN      CDC25.
OS      Saccharomyces kluyveri (yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX      NCBI_TaxID=4934;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=92354938; PubMed=1644315;

```

RA Prigozy T., Gonzales E., Broek D.;  
 RT "Identification and analysis of a DNA fragment from Saccharomyces  
 RT kluyveri that can complement the loss of CDC25 function in  
 RT Saccharomyces cerevisiae.";  
 RL Gene 117:67-72(1992).  
 CC  
 CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. THIS  
 CC PROTEIN POSITIVELY CONTROLS THE LEVEL OF CELLULAR CAMP AT START,  
 CC THE STAGE AT WHICH THE YEAST CELL DIVISION CYCLE IS TRIGGERED.  
 CC -1- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M62964; AAA34479.1; -  
 CC PIR; PC1114; PC1114.  
 CC InterPro: IPR000651; RasGEFN.  
 CC InterPro: IPR001895; RasGEF\_CDC25.  
 CC Pfam; PF00617; RasGEF; 1.  
 CC Pfam; PF00618; RasGEFN; 1.  
 CC SMART; SM00147; RasGEF; 1.  
 CC SMART; SM00229; RasGEFN; 1.  
 CC DR POSITE; PS00720; GDS\_CDC25; 1.  
 CC KW Guanine-nucleotide releasing factor; Cell division; Cell cycle;  
 CC Mtoisis; Transmembrane.  
 CC NON\_TER 1  
 CC FT 959 980 POTENTIAL.  
 CC TRANSMEM 1  
 CC SEQ SEQUENCE 1095 AA; 125605 MW; 7A8B2F9E31A44AC CRC64;  
 Query Match 4.2%; Score 325.5; DB 1; Length 1095;  
 Best Local Similarity 21.6%; Pred. No. 6.1e-08;  
 Matches 18; Conservative 110; Mismatches 240; Indels 225; Gaps 27;  
 QY 434 LEVNGQNFENIQ-LSKAMEILRNTHLSITVKTNL--FVEKELLRLSEKRRNG-----A 485  
 DB 380 LETSADTYKALNTYIDMIEVLEN--LDLTFYVLRNMSHDEKLDRESEELRKHAMTSLA 436  
 QY 486 PHLPKIGDIDKASRYSPDLA---DY-----EYVIGLEKNNKSKAN 525  
 DB 437 STIMEFEIQA---LHDIAIQIMVOGLTLDPEFCAIRDESVCYDRDELSKYQ 492  
 QY 526 TVGGRNKLKILDKTRISIPK-PYNDIGIGSQDSIYGLRQTKIIPALPVSGTLSS 584  
 DB 493 FL---KQEFAEELRNLLISQVEYNDLAFDLA--DQIVREAFKF---SDIGNTMLQ 542  
 QY 585 SNPDLOSRRHRIIDFSATPDLPQVLRVFKADOQ----- 619  
 DB 543 NVEGLIERENILNLYARMMKSDLTALMKGEQKWFEDDFDMASSAEGENENLDFGDAQ 602  
 QY 620 -----RYNI-----SKDTAKVEVYIAIRE 640  
 DB 603 NKSESRDIPWLDSEHYSIYDNKNGIKGTKEALEHNTLSHQSIDLSFNMLMFLFRS 662  
 QY 641 FAVT-----ATPOYSCEVSVPEGY-----IKORRLPDOL----- 672  
 DB 663 IFTTGEFLALVERYNL---YPEEGSYEYINWEEKKPKVIRVYNIMKTLFSQYWT 718  
 QY 673 -----SKLADRIQLS-----GRYLLNNMET---ETLSDDEDAQEL 705  
 DB 719 PSYEPGLDWIGFAQAKSQKISGAVYLLSAIKGRISMKNLKNFPESINSDDOSSST 778  
 QY 706 -----LRESQISLLQSLYEVATQLSKRNFLFRNIEPEYID 743  
 DB 779 TTVPOSSRSSVAPVSSSTTGFRMRKLLDSDIDYAKOLTIKESFLYKISPFECILD 838  
 QY 744 DLF--KLRSKTSKANLKRFEVINOEFVWASLELRETNLKMKIKIKIKIALHREC 801  
 DB 839 RTWGNKTCNNGSKNITEFISNHLTVNVSFMIWVOTDIKRIQLIQFFINVAACHHEL 898

QY 802 KNFNSFAITISGLNLAPEARLRTTWEXLPKNKYELFQDLODDLPFSRNNAKYRNVINSON 861  
 DB 899 NNFSSILATISALYSSPIRYLKRTRMAVPREYKLLLENTLMDSAKNFIRYQLLSIG 958  
 QY 862 LQPIILPFPVIAKDLTEFLHSGDS---KYDGLVNEK---LMIAKEIRHVR----- 909  
 DB 959 -DPPCVFPGVYLSDLTFANGNDPLHRNTVYLVNFGKRYRLLEILKEISVQRSHYKIK 1017  
 QY 910 -----MASVMDPALMEFRTRKKRSLGSLSGSSTNATVYLDVQTGHHK 954  
 DB 1018 RYEDIAFISSLENLPSIRKQYAQSLRNPRTREVSTG---LNTSVNRYNTKNGPENR 1074  
 QY 955 RV-RRSSEFLNAKK 966  
 DB 1075 NTGKRLLKFGKAKK 1087  
 RESULT 4  
 SC25\_YEAST STANDARD; PRT; 1253 AA.  
 ID SC25\_YEAST  
 AC P14771;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE SCD25 protein.  
 GN SCD25 OR SDC25 OR YIL016W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OC NCBI\_taxid=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=M303;  
 RX MEDLINE=91094833; PubMed=1986220;  
 RA Damak F., Boy-Marcotte E., Je Roscouet D., Guilbaud R., Jacquet M.;  
 RT "SDC25, a CDC25-like gene which contains a RAS-activating domain and  
 RL is a dispensable gene of Saccharomyces cerevisiae.";  
 RL Mol. Cell. Biol. 11:202-212(1991).  
 RN [2]  
 RP SEQUENCE OF 668-1253 FROM N.A.  
 RC STRAIN=01136;  
 RX MEDLINE=89306677; PubMed=2545538;  
 RA Boy-Marcotte E., Damak F., Camonis J., Garreau H., Jacquet M.;  
 RT "The C-terminal part of a gene partially homologous to CDC 25 gene  
 RL suppresses the cdc25-5 mutation in Saccharomyces cerevisiae.";  
 RL Gene 77:21-30(1989).  
 RN [3]  
 RP FUNCTION.  
 RX MEDLINE=90260633; PubMed=2188363;  
 RA Crechet J.B., Pouillet P., Mistou M.-Y., Parmegiani A., Camonis J.,  
 RA Boy-Marcotte E., Damak F., Jacquet M.;  
 RT "Enhancement of the GDP-GTP exchange of RAS proteins by the carboxy-  
 RL terminal domain of SCD25.";  
 RL Science 248:866-868(1990).  
 RN [4]  
 RP FUNCTION.  
 RX MEDLINE=91156312; PubMed=2000228;  
 RA Rey I., Schweighofer F., Barlat I., Camonis J., Boy-Marcotte E.,  
 RA Guilbaud R., Jacquet M., Tocque B.;  
 RT "The COOH-domain of the product of the Saccharomyces cerevisiae SCD25  
 RL gene elicits activation of p21-ras proteins in mammalian cells.";  
 RL Oncogene 6:347-349(1991).  
 CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.  
 CC -1- MISCELLANEOUS: SUPPRESSES THE CDC25-5 MUTATION IN YEAST (RESTORES  
 CC CAMP LEVEL) AND HAS SIMILAR FUNCTIONS AS CDC25.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entties requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M26647; AAA16565.1; -

DR PIR: S14177; S14177.

DR SGD: S0003939; SDC25.

DR InterPro: IPR000651; RasGEFN.

DR InterPro: IPR001895; RasGRF\_CDC25.

DR InterPro: IPR001452; SH3.

DR Pfam: PF00018; SH3; 1.

DR Pfam: PF00617; RasGEF; 1.

DR Pfam: PF00618; RasGEFN; 1.

DR SMART: SM00147; RasGEF; 1.

DR SMART: SM00329; RasGEFN; 1.

DR SMART: SM00326; SH3; 1.

DR PROSITE: PS00720; GDS\_CDC25; 1.

DR PROSITE: PS50022; SH3; 1.

KW Guanine-nucleotide releasing factor; Cell cycle; Cell division;

KW SH3 domain.

FT DOMAIN 26 98 SH3.

FT DOMAIN 74 79 POLY-ASN.

FT DOMAIN 434 437 POLY-ARG.

FT VARIANT 584 590 DVVAKFI -> V (IN STRAIN 0L136).

SO SEQUENCE 1253 AA; 144979 MW; 2DE2C9EC27E3E60D CRC64;

Query Match 4.1%; Score 314.5; DB 1; Length 1253;

Best Local Similarity 26.4%; Pred No. 2.4e-07;

Matches 106; Conservative 65; Mismatches 158; Indels 73; Gaps 12;

QY 680 QLSGRYLLKNNMET-----ETLCSDEDAQELLERESOISLLQSTVEVAQOLSMRNF 731

DB 909 EVNOKFKLGNIGEAFAPMKTLDOOIQDHYSGTLYSTE-SILAADPVLFAQTLLHE 967

QY 732 LRRNIEPTIYIDDLER---LRKTSKANLKRREEVINOETFWVASILLETNOLKMKKII 788

DB 968 IYCEITIFPCLOKIMKKNKYKTSYGASPGINEFISFANKLTFNISVYKREADKSKRAKLL 1027

QY 789 KHFIRIALCRECKNFSNFAIISGLNAPVARLRTTWKELPKRYEKLRODDLDLPDSR 848

DB 1028 SHFIPTAECKRFNFSMTAIIISALYSPIYRLEKTMQAVIYDPTDLQSLKMLDPKK 1087

QY 849 NNAKYRNVLNQLPPIIPLEPVIKKDLTLFHEGNSKVDGLVNEFKLIMAKELRHVG 908

DB 1088 NFIVYNELKSLH-SAPCVPFPGVYSLDLTFDTSNPP---DYIVLEHGLKGVHDEKKYIN 1143

QY 909 RRAASVMDPALMRFRKKRKKRWSLGLSGSSTNATVLDVAQTGCHKRVRSSFLNAKKLY 968

DB 1144 -----FNKRSR-----LVDIIQ-----ETIYFKKTH 1164

QY 969 EDAQARKKQYLSN-LELEMDESIQLTSLQCEPATNTLPKNPGDKPKVKSFTSVAR 1027

DB 1165 YDPTKDRIVIECSNLENIPIHEKOYQSLTIEP-----KPKKK---VVPN 1208

QY 1028 ASGQAKQSLPQPOQPPRAHKINQGLQVPAVSLYPSRKVP 1069

DB 1209 SNSNKSQSKSRDDQDTDEKTSFKKD-REPKQLHHTKKKAP 1249

RESULT 5

GNRP\_MOUSE STANDARD: PRT; 1262 AA.

ID GNRP\_MOUSE P27671;

AC P27671;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Guanine nucleotide releasing protein (GNRP) (Ras-specific nucleotide exchange factor CDC25) (CDC25mm).

DE RASGRF1 OR CDC25 OR GRF1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_Taxid=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN-BALB/c;

RC MEDLINE=93010996; PubMed=1396590;

RA Cen H., Lowy D.D.;

RT "Isolation of multiple mouse cDNAs with coding homology to Saccharomyces cerevisiae CDC25: identification of a region related to Bcr, Vav, Dbl and CDC24.";

RT EMBO J. 11:4007-4015(1992).

RN [2]

RP SEQUENCE OF 791-1262 FROM N.A.

RC STRAIN=Swiss; TISSUE=Brain;

RX MEDLINE=92289680; PubMed=1376246;

RA Martegani E., Vanoni M., Zippel R., Cocetti P., Brambilla R., Ferrari C., Sturani E.P., Alberghina L.;

RT "Cloning by functional complementation of a mouse cDNA encoding a homologue of CDC25, a Saccharomyces cerevisiae RAS activator.";

RT EMBO J. 11:2151-2157(1992).

RN [3]

RP SEQUENCE OF 1031-1226 FROM N.A.

RX MEDLINE=92357779; PubMed=1379731;

RA Mei W., Mosteller R.D., Sanyal P., Gonzales E., McKinney D., Dasgupta C., Li P., Liu B.X., Brock D.;

RT "Identification of a mammalian gene structurally and functionally related to the CDC25 gene of Saccharomyces cerevisiae.";

RT Proc. Natl. Acad. Sci. U.S.A. 89:7100-7104(1992).

RL -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.

CC -1- TISSUE SPECIFICITY: BRAIN.

CC -1- SIMILARITY: CONTAINS 2 PH DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entties requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: L20899; AAA02741.1; -

DR EMBL: X59868; CAA42525.1; -

DR PIR: S20730; S20730.

DR PIR: S22693; S22693.

DR MGD: MGI:99694; Rasgrf1.

DR InterPro: IPR001351; GDS\_CDC24.

DR InterPro: IPR000448; IQ\_region.

DR InterPro: IPR001849; PH.

DR InterPro: IPR000651; RasGEFN.

DR InterPro: IPR001895; RasGRF\_CDC25.

DR InterPro: IPR000219; RhogEF.

DR Pfam: PF00169; PH; 2.

DR Pfam: PF00612; IQ; 1.

DR Pfam: PF00617; RasGEF; 1.

DR Pfam: PF00618; RasGEFN; 1.

DR Pfam: PF00621; RhogEF; 1.

DR SMART: SM00015; IQ; 1.

DR SMART: SM00233; PH; 2.

DR SMART: SM00147; RasGEF; 1.

DR SMART: SM00329; RasGEFN; 2.

DR SMART: SM00325; RhogEF; 1.

DR PROSITE: PS50010; DH\_2; 1.

DR PROSITE: PS50041; DH\_1; 1.

DR PROSITE: PS00720; GDS\_CDC25; 1.

DR PROSITE: PS00720; GDS\_CDC25; 1.

DR PROSITE: PS50096; IQ; 1.

DR PROSITE: PS50003; PH\_DOMAIN; 2.

KW Guanine-nucleotide releasing factor; Repeat.

FT DOMAIN 22 130 PH 1.

FT DOMAIN 208 233 IQ.

FT DOMAIN 244 430 DH.

FT DOMAIN 460 588 PH 2.

FT DOMAIN 1025 1259 RAS-GEF.

FT CONFLICT 1033 1033 E -> D (IN REF. 3).

```

SQ SEQUENCE 1262 AA: 144101 MW: 388FE68F7C228DC8 CRC64:
Query Match 4.0%; Score 309; DB 1; Length 1262;
Best Local Similarity 19.8%; Pred. No. 4.4e-07;
Matches 219; Conservative 157; Mismatches 364; Indels 364; Gaps 44;

QY 17 EKHSLPADFTKL-HLSDSLHPQVTHVSSH-----SGCSITSDSGSS-----57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 409 ERNSLDYAKSKLEBSLIMHDEVSETEINIKNLAIEMIEGCEILLDTSQTFVROGSLM 468
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 58 --SLSDIYQATESEAGMDLSGLPETAIVSEDDDEDIRASDPLMSRDIY-----107
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 469 QMSLSEKSSKSRGLSL-----STKGERGOCF-LPSKHLIICTRSGSG 512
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 108 -----RDCKLEDPIDRTD-----DIEQL-----LEFHOHPAANMTMSV 143
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 513 KHLTKNGVSLIDCTILLDEPENLDPAKAGCEIHLERKIGEPKDSLPFVILVAST 572
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 144 RRELCAVM--VFAVVE--RAGTIVLNDGEELDSMSVILNGSEVETYPD--GRAEILCMG 196
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 573 RQKKAAMTSIIQCVDIRCNGMLMNAFEE-----NSKVTVYQMTKSDASLYCDD 622
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 197 NSGCVPTMDKEVYKMGVMTKVDCCQVCIAQODYCRILNOVEKNMOKVEEGEIVWKE 256
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 623 VDIRFSKTNMS-----CKVL-QIR-----640
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 257 HRELDRTGTFKGIHVIKGTSERLTMHLVEHSHVVDPTFIEDFLITYRTFLSSPMYCGKLL 316
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 641 -----VASVERLERLT-----DLRPLSIDLFTFELHSHYKVF--TNMAVVLVDKL 662
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 317 LEWNPDSLDKATRYVLLVNVNH-----FNDEGDPAMTRFLEEFENLERKMGHLR 371
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 683 INIYRKPMASIPARSELLEFSSSHNAKLLYGDAKPSRABR---KSSSPRP-----730
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 372 LNLIAAARAKRRRLMLTKRSREARPLFILLGSEKCGFIVSVDSGSAATAGLKGCD 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 731 -LAIGTSPSPRRKRLSLNIP-----ITGGKALFLA-----760
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 432 QILEVNGQNFENIQLSKAMEILRNTHLSI-----TVKTNLFVEFELTRLSEKRNCA 485
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 761 -SLGSSDSYANI-----HSPISPGKTTLDGKCLMSSSLPKTEPEIDVPA 806
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 486 PHLPKIGDIIKASRYSTPDLAVUEVOYIGLEKYNNKSKANT---VGRNKLKILDKTR 541
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 807 TIPEKGEELSASRKHSHSDYLKESE---DDONHSEDEDTFVSPKSPPTPKSFNRT- 860
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 542 ISILPOKPYDIDIGIOSODSIYGLQTKHIFPALPYSGTFLSSNDPLDLSHRIIDESA 601
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 861 ITEFFEPNNNGILMTTCRDVLVNNNSTISATSAFAIATAGANEGSPKNKEVRR--MSLAN 919
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 602 TPDLPOVLAVFKADOOS--RYTIMISKDTTAKENVIAQATREFAVTATPDQYSLCEVSTP 659
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 920 TG-----FSSDQRINDIKEFVIRRAATNR--VLNVLRHWTKTHQD-----FDY 960
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 660 EGVIKORRL-----PDQLSKLADRIQLSGRYLKNMMEETFLCSDEDAQELLRESQ 710
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 961 DDTLAKYRVICFLEEVNHADPLDQLQ-----ERKAANIIRLTLEETTEQHSMEEYV 1012
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 711 ISLLQST-----VEVATOLSMRNFELFNI-----EPREYIDDLF 746
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1013 LMTGEKVTPEFNHPALEIAEQLTLDHLVKSIPYEEFFGGQMGMAKEYERTPYI---- 1068
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 747 KLRKSTSCANLKRFEVINQETFWVASELIRETNQLRKMLIKHFKIALHCECKNFNS 806
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1069 ---MKTT---KHFNHVS---FIASEIIRNEDISARASAIKWAIVADICCLNINYA 1117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 807 MPAIISGLALAPARLRTTWKELPKNYEKLFQDLFLDPBSRMMAKYRVNLNS--QNLQRP 865
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1118 VLEITSSISIRSAIFRLKTKWLKYSKQTKSLDKLQKLVSD--GRFKNLRESLRCDPP 1174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 866 IIPLEFVINKDLFLEHGDSKV--DGLVNEPKLRMIAKERHVG--RMSVNVNDPLAMR 922
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1175 CVPYLCMYLTLVLFIEGIPNTEDEGLVNSKMRMISHIIREIROFOQTYKIDP----- 1229
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 923 TRKKKMSLSLSOGSTNATVLDVAQTGGHKRVRRSSFLNAKKLYEDNAQMARKVQYLS 982
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1230 -----QPKVIOYLL 1238
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 983 NLELEMDSESLQTLSLQCEPATNT 1006
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1239 DESFMUDEESLYESSLLIERPKLPT 1262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
SOS1_MOUSE STANDARD; PRT: 1319 AA.
ID SOS1_MOUSE 062244;
AC 062245; 062244;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
GN Son of sevenless protein homolog 1 (SOS-1) (mSOS-1).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss; TISSUE=Eye;
RX MEDLINE=92335328; PubMed=1631150;
RA Bowtell D., Fu P., Simon M., Senior P.;
RT Identification of murine homologues of the Drosophila son of
RT sevenless gene: potential activators of ras.";
RN Proc. Natl. Acad. Sci. U.S.A. 89:6511-6515(1992).
[2]
RX STRUCTURE BY NMR OF 415-548
RX MEDLINE=97360234; PubMed=9217262;
RA Koshida S., Kiyawa T., Kim J.-H., Shirozu M., Bowtell D.,
RA Yokoyama S.;
RT "The solution structure of the pleckstrin homology domain of mouse
RT son-of-sevenless 1 (mSOS1).";
RN J. Mol. Biol. 269:579-591(1997).
CC - FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP (BY
CC SIMILARITY).
CC - TISSUE SPECIFICITY: EXPRESSED IN MOST EMBRYONIC AND ADULT TISSUES.
CC - SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.
CC - SIMILARITY: CONTAINS 1 PH DOMAIN.
CC - SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z11574; CAA77662.1; -
CC EMBL: Z11578; CAA77665.1; -
CC PDB: 1PMS: 15-MAY-97.
CC MGD: MGI:98354; SOS1.
DR InterPro: IPR001331; GDS_CDC24.
DR InterPro: IPR004822; Histone_core.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000651; RasGEFN.
DR InterPro: IPR001895; RasGEF_CDC25.
DR InterPro: IPR000219; RhogEF.
DR Pfam: PF00169; PH.1.
DR Pfam: PF00617; RasGEF.1.
DR Pfam: PF00618; RasGEFN.1.
DR Pfam: PF00621; RhogEF.1.
DR SMART: SM00233; PH.1.
DR SMART: SM00147; RasGEF.1.
DR SMART: SM00229; RasGEFN.1.
DR SMART: SM00325; RhogEF.1.
DR PROSITE: PS00741; DH_1; FALSE_NEG.

```



```

DR PROSITE; PSS0010; DH_2; 1.
DR PROSITE; PSS00720; GDS_CDC25; 1.
DR PROSITE; PSS0003; PH_DOMAIN; 1.
KM Guanine-nucleotide releasing factor; 3D-structure.
FT DOMAIN 200 390 DH.
FT DOMAIN 444 548 PH.
FT DOMAIN 777 963 RAS-GEF.
FT DOMAIN 1244 1247 POLY-PRO.
SQ SEQUENCE 1319 AA; 150082 MW; 3286088A5BA0A46 CRC64;

Query Match 4.0%; Score 307.5; DB 1; Length 1319;
Best Local Similarity 19.6%; Pred. No. 5.5e-07;
Matches 224; Conservative 169; Mismatches 409; Indels 339; Gaps 47;

OY 236 NOYEMKMKVVEEGEI-VWVKEHRE--LDRTGKRGHIYIKGSESLTNHLEHSHVNDP 292
DB 240 NDVENFSLIVDHLSVLLGHIEDTVEKTDGSPHPLVGSCEFDLA-----EELAFDP 294
OY 293 --TFIEDFLT--YRTFLSSPMVEGKLL-----EMFNDSPLRDKVTVRYLLVNNHFPND 343
DB 295 YESYARDILRPFHGHFLSOLSKPGALVLIQSIGEGFKE-AVQVLPRLILAVYHCLNH 353
OY 344 FEEDPMATFLFEFENNLREKMGHL-RLN-----IACAKAKRLMTITKRSREA 395
DB 354 FE---LLKOLEEKSEDEQCECKKQAITALLVQSGMEKICSLAKRRRL-----400
OY 396 PLPFIILGSEKGFIFVDSGSKATEAGLKRGDILLEVNGQNPENIQLSKAM-EILR 454
DB 401 -----SESACRYSOOM-----KQOLAKKMMET-OKNIDWEGKDIGCCNEFTM 446
OY 455 NHTLSITVK--TNLFVKELLRLSEKRNCAPLH-----KIGDI 494
DB 447 EGIITVGAKKHEHRIIFLPGLM--ICCKSHGQRLRGASSAFYRLKEKFFMKVQINDK 504
OY 495 KKAASR--SIPDLAVDEOYI-----GLEKVNKSKAMTVGSGNNKLKILDKTRISILPQ 547
DB 505 DDTSEYKHAFFELIKDGNSTYFSAKSAEENKMMALISIQYRSTLERMD-----555
OY 548 KPYNDIGIGSQSDSIVGRQTHIPTALPVSGTLESSNNDLQSHHRL-DESATPDLR 606
DB 556 -----VTVLOEKEEEDMRPLSAVYRFAED--SEENLFEENVOPKAG 597
OY 607 DQYLK--VFKAQDSRYIMISKDTAKEVYIOAIRFAVATATPDQYSLCEVATPEGVI 663
DB 598 IPITKAGTVLKLIERLTHMYADP-----NFVRTFLT--YRSPRQPELLSLI 645
OY 664 KORRLDQSLKADRIQLSG-----RYYLKNMTEFLCS-----D 699
DB 646 ERREIPEPEPTADRIAENGDOPLSAELKRFKREYIOPVQLVLANVCRHWEHNFYDE 705
OY 700 EDQQLRESQ-----710
DB 706 RDADLLQRMEEFGLYGRKAMKMWESTIKIIOKKIARDNGCHNITROSSPTVEMHI 765
OY 711 -----ISLQSLVEVATQLSMNFFELRNIEPTVEYIDLEFLRSK-TSCANLKREE 761
DB 766 SRGHIETFDLILHPIELARQLTLESDDLXRAVQPSSELGVSWTMEKDEKINPNLLKI 825
OY 762 EYVNOCTFWASITLRETNOAKMKIITKFIKIALHCRCKNFNSMPAITSGLNLPVAR 821
DB 826 RHTNTTLTFEKECIVETENLEERVAVVSRIETILQVFOELINNGVLEEVANSNPPYVR 885
OY 822 LRTTWELKNKYEKLFQDLODLDP--SRMAKYRVNVLNSQNQPPIILPFIYIKKDLTF 879
DB 886 LDHTEFQIDPSROKILIEEHELSSEHYKYLAFLRSI-----NPPCVPFEGYLTNLIK 939
OY 880 LHEGNSKVDGLVFEKLEIRIAKEIRHVGSMASVNMDDPALMFTKRRKMSLSLSQSGST 939
DB 940 TREGNP-----EVLRRHGKEL-----INFSKRRVAEITGEIOQYON 976
OY 940 NATVLDVAOTGCHKKRVRRSSFLNAKKLYEDA-----QMAKKVQYLLNLELMEDESIQ 994
DB 977 QPYCLRVER-----DIKRFENILPMGNSMKEETDYILFNKSLFI-----1016

```

```

OY 995 TSLQCEPA-TNTPKPNPGDKK---PVKSE-TSPVAPRAGSOOKAOSLPQOQPPPAKH 1049
DB 1017 -----EPRHKPLRFP--KISTYPLKSPGVPSNPRECTMRH-----PPLQOEP----1060
OY 1050 INQGLVPVAVSLYPSRRKYPVKDLPPEGINSQALKKIILSIEGSLERHKQAEPTSN 1109
DB 1061 -----RKISYSRIP-----ESTESTASA 1079
OY 1110 ASS---QLSSPTSPQSSPPKGYTLAPSGTVDNFSDGSHSEISSRSYSNSSFDPVPS 1166
DB 1080 PNSPRPPLTPPPASGTSNTDVCVSFDS--DHSASPFHSRSASVSISLSKGTDEVYPV 1136
OY 1167 LHDERQORSVSIVETN---LGMGRMERTMIEPDQYSLGSVAPMSGRGLVATFATYISS 1223
DB 1137 PPVPPRRRPPASAPAESPSKINSKHLDSPALPPROPTSKAISP---RISISDRTISID 1192
OY 1224 P 1224
DB 1193 P 1193

RESULT 7
SOS1_HUMAN
ID SOS1_HUMAN STANDARD; PRT; 1333 AA.
AC 007889;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Son of sevenless protein homolog 1 (SOS-1).
GN SOS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eultheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93262494; PubMed=8493579;
RA Charlin P., Camonis J.H., Gale N.W., van Aelst L., Wigler M.H.,
RA Bar-Sagi D.;
RT "Human Sos1: a guanine nucleotide exchange factor for Ras that binds
RT to GRB2.";
RL Science 260:1338-1343(1993).
RN [2]
RP STRUCTURE BY NMR OF 422-551.
RX MEDLINE=98043737; PubMed=9374522;
RA Zheng J., Chen R.H., Corblan-Garcia S., Cahill S.M., Bar-Sagi D.,
RA Cowburn D.;
RT "The solution structure of the pleckstrin homology domain of human
RT SOS1. A possible structural role for the sequential association of
RT diffuse B cell lymphoma and pleckstrin homology domains.";
RL J. Biol. Chem. 272:30340-30344(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 198-551.
RX MEDLINE=99005193; PubMed=9790532;
RA Soisson S.M., Mimmual A.S., Uy M., Bar-Sagi D., Kuriyan J.;
RT "Crystal structure of the Dbl and pleckstrin homology domains from
RT the human Son of sevenless protein.";
RL Cell 95:259-268(1998).
CC 1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.
CC 1- SUBUNIT: INTERACTS WITH GRB2.
CC 1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.
CC 1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC 1- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage debt and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL: L13857; AAA35913.1; -
DR PDB: 1AME; 25-FEB-98.
DR PDB: 1DBH; 23-DEC-98.
DR Genew: HGNC:11187; SOS1.
DR MIM: 182530; -
DR InterPro: IPR001331; GDS_CDC24.
DR InterPro: IPR002119; Histone_H2A.
DR InterPro: IPR004822; Histone_core.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000651; RasGEFN.
DR InterPro: IPR001895; RasGEFN_CDC25.
DR InterPro: IPR000219; RhogEF.
DR Pfam: PF00169; PH. 1.
DR Pfam: PF00617; RasGEFN. 1.
DR Pfam: PF00618; RasGEFN. 1.
DR Pfam: PF00621; RhogEF. 1.
DR SMART: SM00414; H2A. 1.
DR SMART: SM00233; PH. 1.
DR SMART: SM00147; RasGEF. 1.
DR SMART: SM00229; RasGEFN. 1.
DR SMART: SM00325; RhogEF. 1.
DR PROSITE: PS00741; DH. 1; FALSE_NEG.
DR PROSITE: PS50010; DH. 2; 1.
DR PROSITE: PS50020; GDS_CDC25. 1.
DR PROSITE: PS50003; PH_DOMAIN. 1.
KW Guanine-nucleotide releasing factor; 3D-structure.
FT DOMAIN 200 390 DH.
FT DOMAIN 444 548 PH.
FT DOMAIN 777 963 PH.
FT DOMAIN 1258 1261 POLY-PRO.
SQ SEQUENCE 1333 AA; 152463 MM; CGB99C0A11A8DE45 CRC64;

Query Match 3.9%; Score 303.5; DB 1; Length 1333;
Best Local Similarity 19.0%; Pred. No. 8.6e-07;
Matches 221; Conservative 175; Mismatches 397; Indels 371; Gaps 47;

QY 236 NOYEKNQKVEEGEI-VVMKHNRE-LDRGTRKGHIYVKGTSERLTMHLVEE----- 286
DB 240 NDVENIFSRVLDIHLSKLLGHIEDIVEMTDGSPHPLVSGCEDLAELAEDEPYEVSVA 299
QY 287 HSYVDPTFIEDFLTTRTFLLSP-----MEVGKLEMEFNDPSLRDKVTRVYLLAVNNH 340
DB 300 RDLRPEGFHRL-----SQLSKPGALYLOSIGGEKE-----AVGYVLRLLLPYHC 350
QY 341 FNDGDPAMTRLEEFENNLERKMGHL-RLIN-----IACAARRLMTLTGPS 392
DB 351 LHFFE--LLKOLEKSEDEDEKCLKQATLALNVQSGMEKICSLAKRRL----- 400
QY 393 REAPLPIILGSGSEKGFIVDSVDSKATFEGLRGDOILEVNGGNFENIOLSKAM-E 451
DB 401 -----SESACRFYQQM-----KGKOLAIRMNEL-QKNIDGWEGKDIGQCCNE 443
QY 452 ILRNNHLSITVK--TNLFYFKELTRLSEKRNKGNAPLP-----KI 491
DB 444 FMEGSLTRVGAKHERHIFLDGLM--ICKSNHGQRLPGASNAEYRLKEKFEFRKVQI 501
QY 492 GDIKKASRY--STPDLAVDVEQVY-----GLEKYNKKSKANTVGRKKKLKIDTKTISI 544
DB 502 NDDADTNEYNAHEIILKDNSVYFSKASAEKNNMMAALISLOYRSTLEMLDVTM--- 558
QY 545 LPQKPYNDIGIGSQDSDIVGLRQTKHIPTALPVSGFLSSNPDLQSHRILDFSATPD 604
DB 559 -----LQEKKEQMRLPADYVRRAEPP--SEENIIFENNQPK 595
QY 605 LPDQVLR--VFKADQSRIMISKDTTAAKEVYIQAIRFAVATATPDQYSLCEVSVTPEG 661
DB 596 AGPIIAGVIVIKLERLTYHMAVP-----NVRTFLTT-----YRSPCKQKQELLSL 643
QY 662 VIKORLPPDQSLKADIQISG-----RK----- 685
DB 644 IIEFEIPEPEPTADRIATINENGQDPLASIELKRRKEYIQPVOLRVLVNVCNHWVHHFYD 703
```

```
QY 686 -----YLNKME-----TETLCSDPEDAQELLRES----- 709
DB 704 FERDAVLLQMEEEFICTVRGKAKKWEESTYTIQKKIARDGPGHNITFGSSPPTVEM 763
QY 710 -----QISLQSTVEVATOLSRMNFELFRNTEPEYIDDLFKLSRK--TSCANLKR 759
DB 764 HIRPGHIEFTDLTLHPRIEIAQLTLLESDDLRAVQPSLTVGSVTKDEKELNSPNLKL 823
QY 760 FEEVINOETFWVASEILRETNOILRMMKIIFKIALHCRCKENFSMFAISGLNLAVP 819
DB 824 MIHHTNLTFLMEKCIYETENLEBAVAVSRILEIQLVFOELNPNFGLVEVSAWNSSEV 883
QY 820 ARLTWEKLPNKYEKLPQLODLDP--SRNAKRRNVNLNSGNLOPPIIPFPVYKKDL 877
DB 884 YRLDHTPEQIPSRQKKILEAHLSDDHYKKYLAKRSTI-----NPPCVPFGLITLNI 937
QY 878 TFLHEGNDKSVQDLVNFYEKLRIHVGMAVSNMPPALMFTTRKKKWRSLGSLQSG 937
DB 938 LKTEEGNP-----EVLKRRGKEL-----INFSKRRKVAELTGEIQY 974
QY 938 STNATYLDVAQGTGHHKRRVRSFLNAKKLYEDA-----OMARKVOYLSNLELMEDES 992
DB 975 QNOPYCLRVES-----DIKREFENLNPNGSMEMEKEFTDYLFNKSLEIEPRN 1020
QY 993 LQTLSDCEPATNTLPKNPGDK--PYKSE-TSPVAPRAGSQQKASLPQPOQOPPPAH 1048
DB 1021 -----PKPLPRFP--KRYSTPLSPGVPSNPRGTRKH-----PTPLQDEP--- 1060
QY 1049 KINQLOVPVAVSLYPSRKRVVVDLPFGINSPOALKILSLSEGSLEHKKQAEDTIS 1108
DB 1061 -----RKISYSRIP-----ESETESTAS 1078
QY 1109 NASS--QLSSPTSPQSSPRKGYTLAPSGTVNFS--DSGHS----- 1146
DB 1079 APSRPTPLPPPASGASS-----TTDVCASFDSHSPHSSNDVFIQVTLTP 1127
QY 1147 ---EISRSISVSNSSPDSVPVSLHDERORHSVSYETN---LGMGRMRRMTPEPDY 1200
DB 1128 HGRPSASVSSISLTGKIDEPVPPVPPRRRPPSAPRESSPKIMSKHLDSPAIIPROP 1187
QY 1201 SLGSFAPMSEGRGLYATATVYISSP 1224
DB 1188 TSKAYSP-----RYSISDRTSISDP 1207

RESULT 8
CC25_CANAL STANDARD; PRT; 1333 AA.
AC P43069;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cell division control protein 25.
GN CDC25 OR Cdc25.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=792-1;
RX MEDLINE=93238685; PubMed=8477693;
RA Goldberg D., Marbach T., Gross E., Levitzki A., Simchen G.;
RT "A Candida albicans homolog of Cdc25 is functional in Saccharomyces
cerevisiae.";
RL Eur. J. Biochem. 213:195-204(1993).
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GTP BY GTP. THIS
CC PROTEIN POSITIVELY CONTROLS THE LEVEL OF CELLULAR CAMP AT START,
CC THE STAGE AT WHICH THE YEAST CELL DIVISION CYCLE IS TRIGGERED.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M9160; AAA34329.2; -  
DR HSSP: P29354; 1GFC.  
DR InterPro: IPR000651; RASGEFN.  
DR InterPro: IPR001895; RASGRF\_CDC25.  
DR InterPro: IPR001452; SH3.  
DR Pfam: PF00018; SH3; 1.  
DR Pfam: PF00617; RASGEF; 1.  
DR Pfam: PF00618; RASGEFN; 1.  
DR PRINTS: PR00452; SH3DOMAIN.  
DR PRODOM: PD000066; SH3; 1.  
DR SMART: SM00147; RASGEF; 1.  
DR SMART: SM00329; RASGEFN; 1.  
DR SMART: SM00326; SH3; 1.  
DR PROSITE: PS00720; GDS\_CDC25; 1.  
DR PROSITE: PS50002; SH3; 1.  
KW Guanine-nucleotide releasing factor; Cell division; Cell cycle;  
KW Mitosis; SH3 domain.  
FT DOMAIN 30 94 SH3.  
SQ SEQUENCE 1333 AA; 152012 MW; DAF930D6AC192483 CRC64;

Query Match 3.98; Score 303; DB 1; Length 1333;  
Best Local Similarity 19.58; Pred. No. 9.1e-07;  
Matches 214; Conservative 157; Mismatches 339; Indels 390; Gaps 44;

QY 23 ADPTKHLFSLHPOYTHVS-----SSHGCGSTTSDSGSSSLSDIYQAT 66  
DB 432 ASFAQIYINGILSLVHYSLEGFNBSRLFGYDMGKLNKSSSARQSPASSLSTIHQGS 491  
QY 67 E-----SEAGMDLSGLPETAVDSDDEDEIE-----RASDPLMSR 104  
DB 492 DSDTRFAOKLSQDRNSEGNDMNYINQLY--FIDNLRENVNSIVKFLKLSANKKIKNS 549  
QY 105 DIYRDLCKPRIDRTDDDIQLEPMHQL-PAF-----ANNTMSVRELCAMVF 133  
DB 550 DY-----DSSASDDEGDRFDILPQVYRFLVDFNGMNCNPFSTKNTVLNVSGD 602  
QY 154 AVERAGTIVLNGEELDSVILNGSEVETYPDKAELICMGNSEFVSPTMDKEVMKV 213  
DB 603 DLNKRHTKIIIDHSAYDLS-----QIVDKI 629  
QY 214 MTKVDDCOFVCIADQYCRILNQVEKNMQVEEGEIVAKKEHRELDRTGRKGHTVIK 273  
DB 630 ----VQACE-----NILEALDPKVQNTFYNNMLNERNTOI----- 662  
QY 274 GTSERLT-MHLVBEHSVYDPTFIEDLLIYRTFLSSPEVYGGKL--LEMFNDPSLRDKV 329  
DB 663 ---LRLTYKSLVYCSAMVD--LIESF--DFTVFCYSVRHGNADIDEDESEYDPS--- 710  
QY 330 TRVVLNVNHHFDFESD-PAMTRLEEFENNLEREMGSHLLNLTACAAKKRRLMTL 388  
DB 711 ----WGDHYDNLSEFDYVLEDF-----RLKQELHDLVAKIKMTQS---LTL 753  
QY 389 TKPSREAPLFFILLGSEKGFIVDSVDSGSKATEAGLRGQDILEVNOQNENIOLSK 448  
DB 754 EDPE-----VF-----KGLKEEDPL-----FYNRISK 776  
QY 449 -----AMEILRN-----THLSITVKTNLFVEKELLTRLS 478  
DB 777 IPKEKALLSLIKLEQSLFKDGAISLNPDTLLSGYLVIARTKTVLLITQOLI---- 832  
QY 479 EEKRNAGPHLPIGDIKKARSISIPDLAVDEQYIGLEKYNKSKATVGGRNKLKIID 538  
DB 833 EEEETIINATRV---MQDNFDVOLLVBRNNTSSSEKADNS--YVVGGHKK----- 880  
QY 539 KTRISILPQRPYNDIGIGOSDSIVGLRQTKHPTALPVSGTLSSSNPDLQS--HNR 595

DB 881 -----STDVPEWYLEGDEDEYELL-----LDVKGNIKGSKKALVSHLTHHL 920  
QY 596 ILD--FSATPDLDDQVLRVKKADQGSRYIMISKDTAKEVQIAIRFATATAPDQSLC 653  
DB 921 SLDSNFNA-----VF-----LWMFSSNMISGELISLLIARENTE----- 954  
QY 654 EVSATEGV-----IKORRLPDQLSKL-----A 676  
DB 955 -----PEGGLSYEEYNLMWSSKRNPIRLRVINIMKILLKEMWSMYNEPVLRRNLTFNHS 1010  
QY 677 DRIQLSGR-----YLLKNNETTLCSDSD-----AOELLRESOIS---LLQSTV 719  
DB 1011 DOVQVTSIGMLVNYLERLLRGERIYVERDPVJPTKPAFLFTGSSLSKKRPVMDIDV 1070  
QY 720 EVATQLSMRNFELFRNPEPEYIDDLFKLS--KTSQANLKRPEEVNOCFTFWASILR 777  
DB 1071 ELARQTLREFEKLCKYTKTKACLAKWGRKSGSESIDSTIOPFKASNOLNPFYGMILR 1130  
QY 778 ETNOLKRMKTIKHFIKTALCRECKNFNSMFAITSGNLAPVARLPTWEKLPKYEKLF 837  
DB 1131 KADPKRQVQITRFIOYADKCRQYNNFSSMTAITALSYPHRLKKTWEYMADALSNL 1190  
QY 838 QDIQDLFDPSPRNMAKRYRVNLSONLQPIILPFPYIKKDLFTLHEGNSKV--DGLVNF 894  
DB 1191 KNNMKILNNSRNFEYRDVLFKGSE-PCVPFGVYLSDDLTFVYHGNDYLYNRTQGVNF 1249  
QY 895 EKLMTAKELRHVGR-----MASVYNDPALMFR 922  
DB 1250 AKRAKTSSEIVSGIDRFKTTGVNFQVEPEIQFLDAMFEKCPITDEQYQISLNEP----- 1304  
QY 923 TRKKKWRSLGSLSGSTNAT 942  
DB 1305 ---REOPAGASNSNSTTNAT 1321

RESULT 9  
GNRP HUMAN  
ID GNRP HUMAN STANDARD; PRT; 1275 AA.  
AC Q13972;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Guanine nucleotide releasing protein (GNRP) (Ras-specific nucleotide  
DE exchange factor CDC25).  
GN RASGRF1 OR CDC25.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95129875; PubMed=7828890;  
RA Wei W., Das B., Park W., Broek D.;  
RT Cloning and analysis of human cDNAs encoding a 140-kDa brain guanine  
RT nucleotide-exchange factor, Cdc25GEF, which regulates the function of  
RT Ras.";  
RL Gene 151:279-284 (1994).  
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.  
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 IO DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L26584; AAA58417.1; -  
DR Genew; HGNC:9875; RASGRF1.







RA Woodward J., Volckaert G., Aert R., Robben J., Grymompres B.,  
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Wellen-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Landt I., Beck A., Leinach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Medler H., Mambuli R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaie V., Mottier S.,  
RA Galbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpkovski G.V., Usero D., Barrell B.G., Nurse P.,  
RA "The genome sequence of Schizosaccharomyces pombe";  
RL Nature 415:871-880(2002).  
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. THIS  
CC PROTEIN IS ESSENTIAL FOR MATING.  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X53254; CAA37345.1; -  
CC EMBL: AL049559; CAB40184.1; -  
CC EMBL: AL031966; CAA21435.1; -  
CC PIR: S28098; S28098.  
CC InterPro: IPR000651; RasGEFN.  
CC InterPro: IPR001895; RasGRF\_CDC25.  
CC InterPro: IPR001452; SH3.  
CC Pfam: PF00018; SH3; 1.  
CC Pfam: PF00617; RasGEFN; 1.  
CC Pfam: PF00618; RasGEFN; 1.  
CC ProDom: PD000066; SH3; 1.  
CC SMART: SM00147; RasGEFN; 1.  
CC SMART: SM00229; RasGEFN; 1.  
CC SMART: SM00326; SH3; 1.  
CC PROSITE: PS00720; GDS\_CDC25; 1.  
CC PROSITE: PS50002; SH3; 1.  
CC Guanine-nucleotide releasing factor; SH3 domain.  
CC FT DOMAIN 1 60  
CC SEQUENCE 911 AA; 105185 MW; 5F8C12D20C4B753F CRC64;  
SQ  
Query Match 3.6%; Score 277; DB 1; Length 911;  
Best Local Similarity 28.0%; Pred. No. 8, 6e-06;  
Matches 78; Conservative 49; Mismatches 122; Indels 30; Gaps 7;

SOS2\_MOUSE  
ID SOS2\_MOUSE STANDARD; PRT: 1297 AA.  
AC 002384;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Son of sevenless protein homolog 2 (SOS-2) (mSOS-2) (Fragment).  
GN SOS2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Swiss; TISSUE=Eye;  
RX MEDLINE=9235328; PubMed=1631150;  
RA Bowtell D.D., Fu P., Simon M.A., Senior P.V.;  
RT "Identification of murine homologues of the Drosophila son of  
RT sevenless gene: potential activators of ras";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:6511-6515(1992).  
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP (BY  
CC SIMILARITY).  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN MOST EMBRYONIC AND ADULT TISSUES.  
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: 211664; CAA7732.1; -  
CC PIR: S25714; S25714.  
CC HSP: 062245; 1PMS.  
CC MGD: MGI:98355; SOS2.  
CC InterPro: IPR001331; GDS\_CDC24.  
CC InterPro: IPR001849; PH.  
CC InterPro: IPR000651; RasGEFN.  
CC InterPro: IPR001895; RasGRF\_CDC25.  
CC InterPro: IPR000219; RhogEF.  
CC Pfam: PF00169; PH; 1.  
CC Pfam: PF00617; RasGEFN; 1.  
CC Pfam: PF00618; RasGEFN; 1.  
CC Pfam: PF00621; RhogEF; 1.  
CC SMART: SM00233; PH; 1.  
CC SMART: SM00147; RasGEFN; 1.  
CC SMART: SM00229; RasGEFN; 1.  
CC SMART: SM00325; RhogEF; 1.  
CC PROSITE: PS00741; DH\_1; FALSE\_NEG.  
CC PROSITE: PS50010; DH\_2; 1.  
CC PROSITE: PS00720; GDS\_CDC25; 1.  
CC PROSITE: PS50003; PH\_DOMAIN; 1.  
CC Guanine-nucleotide releasing factor.  
CC FT NON\_TER 1 1  
CC FT DOMAIN 164 353 DH.  
CC FT DOMAIN 407 510 PH.  
CC FT DOMAIN 740 926 RAS-GEF.  
CC FT DOMAIN 1145 1148 POLY-PRO.  
CC FT DOMAIN 1164 1173 POLY-PRO.  
CC SEQUENCE 1297 AA; 148846 MW; 3A53F8515BF9D625 CRC64;  
SQ  
Query Match 3.2%; Score 245.5; DB 1; Length 1297;  
Best Local Similarity 18.3%; Pred. No. 0.00042;  
Matches 253; Conservative 188; Mismatches 467; Indels 477; Gaps 60;





```
CC or send an email to license@isb-sib.ch).
-----
DR EMBL AF186779: AAF67280.1: -
DR EMBL AF186780: AAF67281.1: -
DR EMBL AF186798: AAG14400.1: -
DR EMBL AF186781: AAG14400.1: JOINED.
DR EMBL AF186783: AAG14400.1: JOINED.
DR EMBL AF186784: AAG14400.1: JOINED.
DR EMBL AF186785: AAG14400.1: JOINED.
DR EMBL AF186786: AAG14400.1: JOINED.
DR EMBL AF186787: AAG14400.1: JOINED.
DR EMBL AF186788: AAG14400.1: JOINED.
DR EMBL AF186789: AAG14400.1: JOINED.
DR EMBL AF186790: AAG14400.1: JOINED.
DR EMBL AF186791: AAG14400.1: JOINED.
DR EMBL AF186792: AAG14400.1: JOINED.
DR EMBL AF186793: AAG14400.1: JOINED.
DR EMBL AF186794: AAG14400.1: JOINED.
DR EMBL AF186795: AAG14400.1: JOINED.
DR EMBL AF186796: AAG14400.1: JOINED.
DR EMBL AF186797: AAG14400.1: JOINED.
DR EMBL AF186798: AAG14401.1: JOINED.
DR EMBL AF186799: AAG14401.1: JOINED.
DR EMBL AF186785: AAG14401.1: JOINED.
DR EMBL AF186786: AAG14401.1: JOINED.
DR EMBL AF186787: AAG14401.1: JOINED.
DR EMBL AF186788: AAG14401.1: JOINED.
DR EMBL AF186789: AAG14401.1: JOINED.
DR EMBL AF186790: AAG14401.1: JOINED.
DR EMBL AF186791: AAG14401.1: JOINED.
DR EMBL AF186792: AAG14401.1: JOINED.
DR EMBL AF186793: AAG14401.1: JOINED.
DR EMBL AF186794: AAG14401.1: JOINED.
DR EMBL AF186795: AAG14401.1: JOINED.
DR EMBL AF186796: AAG14401.1: JOINED.
DR EMBL AF186797: AAG14401.1: JOINED.
DR EMBL AB023176: BAA76803.1: ALT_INIT.
DR EMBL AL080117: CAB45716.1: -
DR HSSP: 003386: 1LXD.
DR MIM: 60567: -
DR InterPro: IPR000159: RA_domain.
DR InterPro: IPR000651: RasGEFN.
DR InterPro: IPR001895: RasGRF_CDC25.
DR Pfam: PF00617: RasGEF_1.
DR Pfam: PF00618: RasGEFN_1.
DR Pfam: PF00784: RA_1.
DR SMART: SM00314: RA_1.
DR SMART: SM00147: RasGEF_1.
DR SMART: SM00229: RasGEFN_1.
DR PROSITE: PS00720: GDS_CDC25: 1.
KW Guanine-nucleotide releasing factor: Alternative splicing.
FT DOMAIN 236 499 RAS-GEF.
FT DOMAIN 541 613 SER-RICH.
FT DOMAIN 648 735 RBD.
FT VARSPLIC 1 9 MLLINKAKM -> MEYKPYGEPYKSKFKLSTKVESTGHW
FT CONFLICT 15 15 LVEDHYRLWEVLTKEE (IN ISOFORM B).
FT CONFLICT 687 687 W -> G (IN REF. 3).
FT CONFLICT 761 761 N -> Y (IN REF. 3).
SQ SEQUENCE 768 AA: 86700 MW: 5E6FAFGD5E0C2386 CRC64;

Query Match 3.2%; Score 244.5; DB 1: Length 768;
Best Local Similarity 20.7%; Pred. No. 0.00022;
Matches 170; Conservative 112; Mismatches 306; Indels 233; Gaps 31;

QY 605 LPQGVVRVKADQSRVIMSKDT-----TAKRVVQAIREFAVTAPPDYSICE----- 654
DB 77 LVENLITAF-GQNDFTYISIFLSTYRGPASTKVEVLELLDRLGNLTSPN---CEEDGSG 131
QY 655 VSVTPGCVIKO-----RRLPDQ-----LSKLAD-----RIQLS 682
```

```
DB 132 SSSSEKVIKRNALIASILRAMLDQCAEDFREPHFPCLQKLLDYLITRMMPGSDPERRAQN 191
QY 683 GRYYLKNMETER-----LCSDEDAQELLRESQIILLQSTYEVATQLSMRFFELPRNI 736
DB 192 LEQFOKEVEFTDGLPNTISFSLEEELLEGESAEFTCFSEDLVABQILYMAQLFKKV 251
QY 737 EPEYEDDLFKLSK-----TSCANLKRFEVINOETFWVASEL--RETNOQLRMK 786
DB 252 VPRHHCGLMSRRDKKENKHLAFTIRATISQF-----NTLTKCVSYITLGGKELKTQORAK 307
QY 787 IIKHFILALHCRECKNFSMFAIISGLNLAPYARLRTWEKLPNKYEKLFDQLODLFPD 846
DB 308 IIEKWINIAHECRFLKNFSSLRATVSALOSNSYRLKTTAAVPRDMLMFEELDISD 367
QY 847 SRMAKYRVVLSQ-----NLQPL-----IPLPVYIKKL 877
DB 368 HNNHLTSRELLMKEGTSKANLDSVKNOKRTQRRLOLQDKMGVQGVYPTVGTFTDL 427
QY 878 TFLHEGNDGKVD-GLVNFELRMIAKEIRHYGRMASVNDPDMALMFTRRKKMWSLQSLQ 936
DB 428 TMLDTALQDYIEGGLINFERR--REFEYIAQI-----KLLQ 462
QY 937 GSTNATVLDVAQTGCHKRRYRSSFNAKKLYEDAQMARKVKOYLSNLEMDSESLQTL 996
DB 463 SACSNSYCMFPDQ-----KFIFMPQRLTTEES-YAL 494
QY 997 SLOCEPATNTLPKNPDGDKPKVKESTSPVAPRAGSQOKAOSLPQOQPPPAHINQLOY 1056
DB 495 SCEIEAABASTTSPKPRKSMVRLSLF--LQSDMITSTPT--TKRQPKTAGSGSGESM 551
QY 1057 PAVSL--YPSRKVPVVKDPPFGINSPOAKKILSLSEEGSLERHKKQAEPTISNASSQ 1113
DB 552 DSVSVSCESNHSSEAEBSGTT--MDTPPEQKKLESSSSCSGSIH---SMDTNSGMS 606
QY 1114 LSSPPTSPQSSPRKGYTLAPSGTVDNFSDGSEISRSIVSNSSPVSFLHDERQ 1173
DB 607 LNPPLSSPPSC-----NNPKIHKRSVSVSISSTVLPVYNOQNE 648
QY 1174 RHSVSI-VETNLGMKMERRTMLEPDQYSLGSIAPMSGGLYATATVYSSPTEELSOD 1232
DB 649 TCIIIRISVEDNNG-----NMKRSIMLTSQDKTPAVIOR 681
QY 1233 QCGRASLDAAADSGRGSWTSSSGSHDNIQTIQHRSMETLPFGHTHFDEYGGDPAGL 1289
DB 682 AMLKHNLD-----SDPAEYELVQVISEDKEI-VIP-----DSANFYAM 720
QY 1290 ASSSHMDQIMFSDHSTKYNRONOSRES--LEQAOSRASMAS 1328
DB 721 NSQVNFDFILRKKNMSMEQYKLSRSLTLPRTAKKGCSN 761

RESULT 15
SOS2_HUMAN STANDARD; PRT; 1332 AA.
ID AC 007890: Q15503:
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Son of sevenless protein homolog 2 (SOS-2).
GN SOS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93262494; PubMed=8493579;
RA Chardin P., Camonis J.H., Gale N.W., van Aelst L., Wigler M.H.,
RA Bar-Sagi D.;
RT "Human Sos1: a guanine nucleotide exchange factor for Ras that binds
RT to GRB2.";
RL Science 260:1338-1343(1993).
```

[2]  
RN SEQUENCE OF 724-1296 FROM N.A.  
RP TISSUE=Placenta;  
RC Fath I., Apioa F., Dutrillaux B., Tocque B.;  
RL Submitted (MAR-1996) to the EMBL/GenBank/DBD databases.  
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP (BY  
CC SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: L33858; AAA35914.1; -;  
CC EMBL: L20686; AAA91852.1; -;  
CC HSSP: Q62245; 1PMS.  
CC Genew: HGNC:11188; SOS2.  
CC MIM: 601247; -;  
CC InterPro: IPR001331; GDS\_CDC24.  
CC InterPro: IPR001849; PH.  
CC InterPro: IPR000651; RasGEFN.  
CC InterPro: IPR001895; RasGEF\_CDC25.  
CC InterPro: IPR000219; RhogEF.  
CC Pfam: PF00169; PH; 1.  
CC Pfam: PF00617; RasGEF; 1.  
CC Pfam: PF00618; RasGEFN; 1.  
CC Pfam: PF00621; RhogEF; 1.  
CC SMART: SM00233; PH; 1.  
CC SMART: SM00143; RasGEF; 1.  
CC SMART: SM00229; RasGEFN; 1.  
CC SMART: SM00325; RhogEF; 1.  
CC PROSITE: PS00741; DH\_1; FALSE\_NEG.  
CC PROSITE: PS50010; DH\_2; 1.  
CC PROSITE: PS00720; GDS\_CDC25; 1.  
CC PROSITE: PS50003; PH\_DOMAIN; 1.  
CC KW Guanine-nucleotide releasing factor.  
CC -----  
FT DOMAIN 198 388  
FT DOMAIN 442 546  
FT DOMAIN 775 961  
FT DOMAIN 755 758  
FT DOMAIN 1180 1183  
FT DOMAIN 1203 1208  
FT DOMAIN 778 778  
FT CONFLICT 861 861  
FT CONFLICT 948 948  
FT CONFLICT 999 999  
FT CONFLICT 1032 1032  
FT CONFLICT 1042 1042  
FT CONFLICT 1112 1114  
FT CONFLICT 1159 1159  
FT CONFLICT 1296 1296  
SQ SEQUENCE 1332 AA; 153030 MW; EFABA952FA85AC9 CRC64;  
Query Match 3.18; Score 242.5; DB 1; Length 1332;  
Best Local Similarity 17.08; Pred. No. 0.00061;  
Matches 220; Conservative 135; Mismatches 322; Indels 615; Gaps 50;

Db 458 HIFLFDGLMISCKPNHGTRLPQYTSAEYRLKEKFMKRIQCDKEDNCEHKAHELYSK 517  
QY 229 QDYCRLL---NOVEKN-----MOKVEEGEIVVYKREHL----- 260  
Db 518 DENSIIFAAKSAEKKNNMAALISLHYRSTLDMDSVLKKEENQPLRLPSEYRYEV 577  
QY 261 -----DRTGRKHIVIKGTS-----EELTMHLYEESHVDPPIEDFLTYRTF 305  
Db 578 KQSEENVIVEDNQSRSGLPIIKGTIVVKLIERLYHM-----YADPNVRFRFLTYRSF 632  
QY 306 LSSPMVEVGRKLEWFN---DPSLRDK-----VTRVY 333  
Db 633 -CKPQELLILIRFRIPEEPDADKLAIEKEGEPIADLKRFRKEYQVQYQLRYLNF 691  
QY 334 LHMVNNHNDPEEDPAMTRLEFEFNNLREKMGHLRLNTACAKAKRRLMTLTKPSR 393  
Db 692 RHVVDHYHDEFDLLELLELSEFISVR-----GKAKK----- 726  
QY 394 EADLPFLLGSEKGFGEIVDSVDSGSKATEAGLRGQDILEVNGQNFENIOLSKAMEIL 453  
Db 727 -----WYEST----- 731  
QY 454 RNNTHLSTVYKTNLEFVKELLRLSEKRNAGAPHLKIDIKKASKYSIPDLAVDVEOYI 513  
Db 732 -----AKIIRKKQAOANGVSH----- 748  
QY 514 GLEKYNKSKANTVGGRNKLKILDKTRISILPQKRYNDIGIGOSDDSYVGLRQTKHIP 573  
Db 749 -----NITFESP----- 755  
QY 574 TALPVSGTLSSNPDLLOSRRILDEFSAPPDLPDQVLRVFKADQDSRYIMISKDTAKEV 633  
Db 756 -----PPIEMH-----ISK----- 765  
QY 634 VIOAIRFEAVTAPPDQYSLCEVSTPEGVYKORRLPDQLSKLADRIQLSGRYLKNNMET 693  
Db 766 -----PGQF----- 769  
QY 694 ETLCSDEDAQELRESQISLLOLSTVEVATQLSMRNEFLERNIEPTYEDLFLKRSK-T 752  
Db 770 ET-----FDLMTLDPIELIAQQLLLESDDLTKRQPSBELVGSVTKRDKEL 814  
QY 753 SCANLKRFEVINOETFWVASEILRETNLQKRMIKIKHFKIALHCRCKENSMFAIIS 812  
Db 815 NSPMLKMIHTTNLTLMFEKCIYEAEENFEERVAVLSRIEILQVQDNLNENGVALEIYS 874  
QY 813 GNLIAVYARLRTWEKLPNKYTEKLFODLDLPSPSNMAYKRYVNLVSQNLQPIILFPV 872  
Db 875 AVNSVSYYRLDHTFEALQERRKRIIDEAVAL--SDHFKKY--LVKLKSTINPCVPEFGI 930  
QY 873 IKKDLTFHEGNS---KVDGLVNFELKRLMAK---EIRHVG-----BMA5V 913  
Db 931 YLNIILTEGNNDFLKRKGDILNFSKRRKVALEITGEIQOYONOPYCLRIEPPDMRFE 990  
QY 914 NMDP-----ALMEFTRKKW-----RSLGSLG-----QGSTNA 941  
Db 991 NLNMGASAEKEFDYLLFNKSLIEEPKNCQPPRPKSPFLSKGIRPNTGHSSTG 1050  
QY 942 TYLDVATQGGHKRKYRS---SFLNAKKLYEDAQAARKYKQYLSNLELMEDESLQTLIS 997  
Db 1051 TL-----RGHPPLEREKPKISF-----SRIATELE-----STVS 1081  
QY 998 LOCEPATNTLPKNGDCKPKPKSET-----SRIATELE-----STVS 1081  
Db 1082 APTSPNTPTSP-----PVSASDLSVFLVDVNLSSGNSIFAVYLLPHSKSFSSCG 1134  
QY 1023 -----PVADR-----AGSQKASLPO-PQOQPPRAKINQGLQVPA 1058  
Db 1135 SLNKLSEPLPLPPLPRKKFHDHANSKGNMKSDDDPAPAIIPROPPPP-KVAPRYVPPT 1193  
QY 1059 -----VSLYPSKKKYVPVNDLPP-----FGINSQALKKILSLSESGSLERHKKQA 1103  
Db 1194 GAFDGPLHSPPPPPPRDLDPDTPPYVPLRPPEHINCPFNLDP---PVLGHILRSDWL 1249

Wed Feb 19 16:11:35 2003

us-09-911-826a-2.rsp

Page 17

```
Qy 1104 EDTISNASSQLSSPPTSPOSS-PRKGYTLAPS 1134
      | | | : | : | : | | : | : |
Db 1250 RD-ISTCNSPSTPPSTPSPRVPRRCYVLSSS 1280
```

Search completed: February 18, 2003, 21:24:14  
Job time : 61 secs

---

**BLANK PAGE**